For Official Use **ENV/JM/TG/M(2002)2**

Organisation de Coopération et de Développement Economiques
Organisation for Economic Co-operation and Development **30-Sep-2002**

English - Or. English

ENVIRONMENT DIRECTORATE
JOINT MEETING OF THE CHEMICALS COMMITTEE AND
THE WORKING PARTY ON CHEMICALS, PESTICIDES AND BIOTECHNOLOGY

ENV/JM/TG/M(2002)2
For Official Use

Test Guidelines Programme

REPORT OF THE STOCKHOLM CONFERENCE ON VALIDATION AND REGULATORY ACCEPTANCE OF NEW AND UPDATED METHODS IN HAZARD ASSESSMENT

Mr. Herman Koëter
Tel: + 33 (1) 45 24 98 44, Fax: (33) 1 45 24 16 74 , E-mail: Herman.Koeter@oecd.org

JT00132299

English - Or. English

EXECUTIVE CONFERENCE SUMMARY

An OECD Conference on Validation and Regulatory Acceptance of New and Updated Methods in Hazard Assessment was held in Stockholm, Sweden, 6-8 March 2002. The stated purpose of the conference was to develop, and achieve consensus on, practical guidance on principles and processes for the validation and acceptance of animal and non-animal test methods for regulatory hazard assessment purposes. This consensus guidance would be used to revise the draft OECD Guidance Document (No. 34) on "The Development, Validation and Regulatory Acceptance of New and Updated Test Methods in Hazard Assessment"".

The conference agreed on practical guidance for assessing the reliability and relevance of animal and non-animal testing methods used to assess the safety of chemical substances and products.

The agreement includes detailed guidance on a benchmark for the quality of testing methods. This Test Data Interpretation Prediction (formerly referred to as "Prediction model") will help scientists and regulators to convert results from tests into a prediction of the hazard so they can make regulatory decisions. Animal, in vitro and ethical human tests are the basis for predicting the hazards of chemicals on human health and the environment. These tests should be able to predict expected effects on humans or the environment in order for the regulators to be able to rely on their outcomes. The agreed Data Interpretation Prediction will provide for each test a scientific description explaining what the effects observed in the test would predict for human health or the environment. To this end, test methods need to be frequently updated to include the latest scientific developments while consideration is given to the welfare of the test animals

The Conference further agreed on practical guidance for the application of the 3 R's principles (Replacement, Reduction and Refinement of animal tests) in method development. They also agreed on an independent review process to confirm the quality of each validation study and facilitate the regulatory use of new methods in hazard assessment. Although independent peer reviews are common practice for acceptance of publications in the scientific literature, structured and transparent independent peer reviews are currently only marginally applied in the acceptance process of new test methods for hazard assessment.

The Conference strongly recommended that separate Workshops or expert meetings be held on the use of human data and the Data Interpretation Prediction, respectively, to develop more detailed guidance on these issues.

TABLE OF CONTENTS

EXECUTIVE CONFERENCE SUMMARY 2

BACKGROUND 5

CONFERENCE PURPOSE AND OBJECTIVES 5

CONFERENCE STRUCTURE AND PARTICIPANTS 6

Breakout Groups 6

CONFERENCE DOCUMENTS 6

Discussion Documents and Supporting Discussion Documents 6
Discussion Documents: 7
Supporting Discussion Documents: 7
Conference Background Documents 7

CONFERENCE INTRODUCTION AND SETTING THE SCENE 8

SUMMARY REPORT OF THE CONFERENCE DISCUSSIONS AND RECOMMENDATIONS 8

The Solna Document and Principles 8
General Comments and Organisation of the Guidance Document 9
The Title of the Document 10
Data Interpretation Procedure (formerly known as the Prediction Model) 11
The Validation Management Group, Peer Consultation, and Peer Review 12
The Validation Management Group 12
Expert Consultation by VMG 13
Peer Review 14
Test Development 15
The Test Sponsor 16
The Validation Process 16
Prevalidation 17
Validation 17
Validation of Test batteries 18
Validation of QSAR Systems 18
“Catch-Up” Validation 18
Retrospective validation 19
Transparency 20
Reference Chemicals and Data 20
Reference Chemicals. 20
Reference Data 20
Animal Welfare 22
GLP Compliance 22
Patented Methods 23
Regulatory Acceptance 23
Financial Sponsorship or Support of Validation Studies 25

GENERAL EDITORIAL AND FORMATTING ISSUES 25
Figures and Tables 25
Definitions and Glossary.....26
Editorial and Other Specific Recommendations for the Draft Guidance Document 26
Other Areas of Concern or Comment 27
GENERAL CONCLUSIONS AND RECOMMENDATIONS 27
ANNEX 1 30
ANNEX 2 ..... AVAILABLE TO GOVERNMENT OFFICIALS ONLY
ANNEX 3 35
ANNEX 4 36
ANNEX 5 -10 51
ANNEX 11 117

BACKGROUND

1. The 13th WNT agreed that it was timely to arrange for a follow-up meeting to the 1996 Solna Workshop on Validation and Regulatory Acceptance of New and Alternative test Methods. A follow-up meeting was considered necessary to provide further guidance on the interpretation and application of the Solna principles, taking into account current concerns with respect to children's health, endocrine disrupters and other health concerns and, at the same time, ensuring that the use of animals would be limited to the extent possible and that all test methods used are relevant and reliable. Sweden and the USA offered assistance with the arrangement of this Solna follow-up meeting. It was agreed that preparatory meeting(s) would be hosted by the US and that the Conference would be held in Sweden.

CONFERENCE PURPOSE AND OBJECTIVES

2. The OECD Conference on Validation and Regulatory Acceptance of New and Updated Methods in Hazard Assessment was held in Stockholm, Sweden, 6-8 March 2002. The stated purpose of the conference was to develop, and achieve consensus on, practical guidance on principles and processes for the validation and acceptance of animal and non-animal test methods for regulatory hazard assessment purposes. This consensus guidance would be used to revise the draft OECD Guidance Document (No. 34) on "The Development, Validation and Regulatory Acceptance of New and Updated Test Methods in Hazard Assessment".

3. The specific Conference objectives, as presented to the participants, were:

(i) Provide practical guidance on how to adequately address established validation principles and criteria;

(ii) Provide practical guidance on the conduct and management of the validation process;

(iii) Provide practical guidance on how to adequately address established principles and criteria for regulatory acceptance of validated test methods including the submission of information to support their validity;

(iv) Provide practical guidance on the process for independent peer review, regulatory consideration and implementation of new and updated test methods.

4. A Steering Committee of individuals nominated by Member countries was established to develop the agenda and structure of the Conference and to advise on documents, presentations and invited speakers, chairpersons, and rapporteurs. The Steering Committee met once in Washington, DC US and had numerous teleconferences to prepare and arrange for the Conference. The Steering Committee members were:

David Blakey, Canada
Toini Berzins, Sweden
Julia Fentem, UK
Tohru Inoue, Japan
Hiroshi Ono, Japan (alternate to T. Inoue)
Horst Spielmann, Germany
William Stokes, USA/ICCVAM
Gary Timm, USA
Atsuya Tagaki, Japan
Bo Wahlström, Sweden
Andrew Worth, EC/ECVAM
Yoshikuni Yakabe, Japan

CONFERENCE STRUCTURE AND PARTICIPANTS

5. The Conference was structured as an alternation of plenary and break-out group sessions. Four break-out groups each addressed one of the four specific Conference objectives (see paragraph 3) in a series of sessions. These sessions were alternated by plenary sessions, ranging from a general introductory session via sessions discussing break-out group progress reports to sessions focusing on overall consensus building and conclusions/recommendations. Details of the Conference Agenda are provided in Annex 1.

6. Each Break-out group had two co-chairs and two co-rapporteurs. Chairs and Rapporteurs were selected by the Steering Committee based on knowledge of and experience with method development and validation, regulatory hazard and risk assessment, and chairing and/or reporting of discussion meetings.

7. The conference participants included close to 100 representatives from 13 Member countries, the European Commission, BIAC, and the newly formed International Council for Animal Protection in OECD Programs (ICAPO). The participants, and their affiliations are in Annex 2.

Breakout Groups

8. Breakout Group (BG) assignments were made prior to the meeting. Some participants, in response to a request by the Secretariat, asked to be assigned to specific BGs. Individuals who did not express a preference were assigned by the Secretariat. At the meeting, a number of individuals who had not requested specific BG assignments asked if they could participate in a BG to which they were not assigned. A number of individuals -- Steering Committee members and others-- were not assigned to specific BGs, but circulated among the BGs as participants and observers. In addition, OECD Secretariat staff circulated among the BGs as observers and to respond to any questions the groups may have had regarding the purpose of the conference and OECD's role in validation studies. As a result, the listing of BG assignments (Annex 3) does not totally accurately reflect the actual makeup of the BGs during the meeting. The charges to the BGs, and the specific questions they were asked to address are in Annex 4.

9. In addition to the formal charges to the BGs, the co-chairs and rapporteurs of the BGs were reminded that the underlying purpose of conference was to assist the OECD Secretariat in developing practical guidance and recommendations to improve Guidance Document No. 34.

CONFERENCE DOCUMENTS

Discussion Documents and Supporting Discussion Documents

10. The number of (Supporting) Discussion Documents was limited to those that were considered to form the basis of, or provide material/text for, the Guidance Document referred to in the "Purpose of the Conference" (see paragraph 2). The following documents were considered as useful for the various discussions as (Supporting) Discussion Documents:

Discussion Documents:

- OECD Draft Guidance Document No.34: Development, Validation and Regulatory Acceptance of New and Updated Internationally Acceptable Test Methods in Hazard Assessment;
- Compilation of comments from Member countries' experts on Draft Guidance Document No.34;

Supporting Discussion Documents:

- OECD Document ENV/JM/TG(2001)5: Validation Issues: Current Practices and Issues for Consideration. Annexed to this document are comments from US.EPA, ECVAM and ICCVAM;
- OECD Document ENV/MC/CHEM/TG(96)9: Final report of the OECD Workshop on Harmonisation of Validation and Acceptance Criteria for Alternative Toxicological Test Methods;
- ICCVAM Document: Validation and Regulatory Acceptance of Toxicological Test Methods, NIH Publication No. 97-3981;
- ICCVAM Document: Evaluation of the Validation Status of Toxicological Methods; General Guidelines for Submissions to ICCVAM (Revised, October 1999);
- ECVAM Workshop Report No.5: Practical Aspects of the Validation of Toxicological Test Procedures, ATLA 23, 129-147, 1995;
- ECVAM Prevalidation Task Force, Report No.1: The Role of Prevalidation in the Development, Validation and Acceptance of Alternative Tests, ATLA 23, 211-217, 1995;
- The Role of ECVAM in Promoting the Regulatory Acceptance of Alternative Methods in the European Union (A.P. Worth and M. Balls, ATLA 29, 525-535,2001).

11. These (Supporting) Discussion Documents were all made available electronically to all conference participants prior to the conference.

<u>Conference Background Documents</u>

12. In addition to the (Supporting) Discussion Documents a considerable number of background documents were received from Member countries, ECVAM, ICCVAM and other stakeholders on:

- principles of method development and validation,
- national and international validation studies,
- independent peer review processes, and
- hazard assessment approaches and animal welfare considerations.

13. These background documents were considered as useful for the discussions (of the Break-out Groups) as they provided: (i) detailed explanations of approaches, viewpoints and principles, (ii) justifications for particular approaches, and (iii) examples of various validation studies that are extremely useful as "lessons learned" or "success stories". However, for practical reasons these background documents were made only available (as paper copies) during the conference.

CONFERENCE INTRODUCTION AND SETTING THE SCENE

14. The Conference was opened by Ms. Ethel Forsberg, Director General of the Swedish National Chemicals Inspectorate (KEMI). In her welcome she emphasised the need for internationally accepted alternative methods and, consequently, the need to reach full agreement on criteria and principles to evaluate the reliability and relevance of any new test.

15. Herman Koëter of the OECD Secretariat informed the Conference of the historical perspective of the issues to be addressed and provided further details of the Conference objectives and the role of the participants. He emphasised that OECD's first priority in hazard assessment is human health and environmental safety. Furthermore he urged the meeting not to re-invent the already agreed principles and criteria for validation and regulatory acceptance of new tests but instead provide practical guidance on how to produce evidence that a test method for hazard assessment is relevant and reliable and ready for use in regulatory assessment and how to present the evidence to the regulatory authorities in a way that facilitates the acceptance and use. A copy of Mr. Koëter's presentation is attached to this report as Annex 5.

16. The Secretariat's introductory presentation was followed by presentations from Japan (Mr. Tohru Inoue) and North America (Ms. Susan Hazen) offering the participants their respective perspectives on the conference purpose and objectives. Their presentations are attached as Annex 6 and 7, respectively.

17. Following these Member country presentations, the EU ECVAM's representative Mr. Michael Balls, the US ICCVAM representative Mr. William Stokes and the industry BIAC representative. Mr. Mark Chamberlin shared their views on the Conference, the challenges and issues relevant to their respective centers and industry and their preferred approaches for validation and acceptance of new methods. Their presentations are attached as Annex 8, 9 and 10, respectively.

SUMMARY REPORT OF THE CONFERENCE DISCUSSIONS AND RECOMMENDATIONS

18. The following summarises the discussions in the Breakout Groups, the daily Breakout Group Reports to the plenary sessions, and the discussions in the plenary sessions. The ordering of the sections in this report does not strictly follow the order in which the issues were raised and discussed during the Conference. Furthermore, the ordering is not intended to direct the ordering or nomenclature of the sections in the Guidance Document. Throughout this report, references are made to sections or paragraphs in the draft Guidance Document. Recommendations for specific changes or topics to be included in the Guidance Document are highlighted in boxes. The Summary Reports and Statements of the various Breakout Groups are attached to this Report as Annex 11, for Breakout Group 1,2,3 and 4, successively.

The Solna Document and Principles

19. The Conference generally endorsed the Solna document and the validation principles it presents (the "Solna principles"). The emphasis of the discussions was directed towards clarifying the concepts and definitions presented in the document, and ensuring that it was written so as to be comprehensive for all types of tests needing validation, and that it could be usable by different countries and organisations that may have different regulatory needs or philosophies.

Recommendation 1:
It was recommended that because Guidance Document No. 34 is based on the Solna principles, a summary of these principles be prepared and appended to the Guidance Document.

General Comments and Organisation of the Guidance Document

20. Because this Guidance Document is being developed to support the OECD Test Guideline Programme it should include a brief description of the programme, possibly in an Annex. This description should include the process of the development and acceptance of a new or revised Test Guideline, and the position of test validation in this process.

21. Test validation should be a hypothesis-driven process that uses the Solna principles. The basic validation principles (e.g., OECD-Solna/ECVAM/ICCVAM) are appropriate and workable, and should be simply stated. The purpose of the Guidance Document is to provide a broad, generic document which will guide individuals and organisations through the test validation process, and provide examples from validation studies that have been carried out by different organisations applying a variety of approaches.

22. The Conference agreed that the following types of tests and procedures needing validation should be addressed in the Guidance Document. The list is not to be regarded as complete or all inclusive and, therefore, not to be seen as limiting, but only to exemplify the considerable range:

- Human tests
- *In vivo* animal tests
- *In vitro* tests
- SAR procedures
- Genomics/proteomics and other novel techniques
- Multiple endpoint studies
- Test batteries
- Tiered test systems
- Ecotoxicity tests
- Statistical methods
- New tests
- Substitute tests
- Modification of endpoints in existing tests
- Existing test methods needing retrospective validation

23. There was a general consensus that the draft Guidance Document was somewhat too prescriptive and detailed in several places, and should be simplified. Conversely, it was also recommended that there were areas of the document that were too general and that more detail and specificity were needed. Despite this apparent contradiction, there was consensus that the Guidance Document should provide general guidance and a structure to use for test validation, and not be a checklist of details.

24. It was further agreed that the planning and conduct of a validation study should be undertaken on a case-by-case basis to take into consideration the individual components of test validation that need to be included in the study, based on the nature of the test, its intended use, and the nature and extent of relevant, prior validation studies.

25. The important principle behind a validation study is to ensure that the proposed test method will be able to predict the activity of chemicals for the endpoint of concern with sufficient accuracy. The validation process can be used for new tests and, if deemed necessary, for tests that have been accepted by convention, but have not been formally validated. With respect to the latter, the meeting emphasised that in case validation is not considered, a written justification should be available.

26. Appropriately qualified groups or organisations should be involved in the planning and conduct of validation studies, regardless of whether they are formal centers for validation (e.g., ECVAM) or for review of validation studies (e.g. ICCVAM). Validation studies could also be conducted by research organisations/centers and industry laboratories.

27. The draft Guidance Document currently does not address specifically the validation of ecotoxicology tests.

Recommendation 2:
The introductory paragraphs of the Guidance Document should be rewritten to note that validation studies can range from small scale (i.e., run by individuals or small groups of investigators) to large-scale, multi-national programmes. These should also be reference to a description of the Test Guidelines Programme and the position of test development and validation in this process.

Recommendation 3:
The Guidance Document should not be a checklist of details to be considered but, instead, provide general guidance and a structure for test validation.

Recommendation 4:
The Guidance Document should include a section on the specific aspects of validation of ecotoxicology tests.

The Title of the Document

28. A number of discussions included suggestions for changes in the title of the Guidance Document because the word "development" is confusing. In OECD "development" means the development of a Test Guideline which is a well-defined process that starts with a well-described test method, includes validation and ends with adoption of the method by OECD Council. In other circles "development" means only the first part: the scientific development of a method i.e., the phase that precedes prevalidation.

Recommendation 5:
Because it does not explicitly address method development, a title change to "The Validation for Regulatory Acceptance of New and Updated Internationally Acceptable Test Methods for Hazard Assessment" was recommended.

Data Interpretation Procedure (formerly known as the Prediction Model)

29. There were extensive discussions in the breakout groups and in the plenary sessions regarding the need for a formal data interpretation procedure, its definition and scope, and whether the term "prediction model" was the most appropriate term to use. It defines the steps that must be taken to convert the results from a toxicity test into a prediction of toxicity that may be used in a safety assessment. It was recognised that the concept is widely accepted by the community of scientists who develop and validate in vitro toxicity tests. Conference members indicated that an in vitro test is useless if it does not have an adequate data interpretation procedure. Such a procedure or concept is necessary because it provides a link between an in vitro test result and a prediction of toxicity that is ultimately used in the safety assessment process. Participants of the Conference were familiar with this concept and recognised its importance, however, a considerable number of participants found that the "prediction model" confusing and limited to certain types of studies, others believed that the term was unambiguous and covered all types of validation studies.

30. It was noted that the term "prediction model" was originally developed for the validation of in vitro tests designed to predict toxicity in vivo. Some members of the conference expressed concern that the term, algorithm, implies a mathematical model which is often not appropriate. The developers of the concept, who participated in the Conference, acknowledged this concern, but pointed out that the original definition of the concept does provide for prediction models that are not necessarily mathematical. The conference agreed that the term 'data interpretation procedure' is a more general term, (though full consensus was not reached on the issue,) that could apply to both in vitro and in vitro studies, and might be less confusing. Therefore the term 'data interpretation procedure' was adopted to be used instead of the term 'prediction model' in the Guidance Document.

31. A number of examples of data interpretation procedures were presented, which included both mathematical and statistical models, and nonmathematical descriptions of the relationship between the anticipated test results and the effect of interest. A single test can have different data interpretation procedures, depending on the various test uses being modelled.

32. The Conference agreed that, regardless of the terminology used, illustrative examples of quantitative and qualitative data interpretation procedures be incorporated into the Guidance Document.

33. It was noted that by modifying the data interpretation procedure following (pre) validation work to fit the data, one may create a self-fulfilling prophesy as a result of circular reasoning. The Guidance Document should provide more guidance than is currently included to explain this issue.

34. It was considered necessary to provide additional guidance on how to describe data interpretation procedures for new assays that do not have any, or sufficient, relevant reference data. Examples from previous validation studies should be included as case studies.

Recommendation 6:
The phrase 'Data Interpretation Procedure', is preferred to 'prediction model' because it is broader and less ambigiuous. Furthermore, the following was considered as an adequate description of the data interpretation procedure, and therefore, should be included in the Guidance Document:

> Data Interpretation Procedure (DIP) defines the relationship between the results of the test and the toxicological concern. In other words it defines the steps to be taken to convert results from a test method into a prediction of hazard for the species of concern that is useful for making decisions. The relevance of the data interpretation procedure should be assumed as a part of the validation study.

Recommendation 7:
More guidance should be provided on the issue of modification/adaptation of the data interpretation procedure based on results of the first phase of a validation study.Examples of different types (i.e., mathematical; descriptive) of data interpretation procedures should be included in the Guidance Document.

The Validation Management Group, Peer Consultation, and Peer Review

35. The Conference agreed that there should be a clear distinction in concept and process between the validation of a method and the peer review of that validation.

36. The different stages from test development through peer review of the validation study, including the VMG and peer consultation, have been defined and differentiated in Table 1, which was developed by Breakout Group 4 (see Summary Report of Breakout Group 4 in Annex 11). However, time did not allow to discuss the scheme in plenary and reach consensus on it.

The Validation Management Group

37. A number of breakout groups extensively discussed the various aspects of the validation management group. The various views and opinions were also discussed at length in plenary sessions and resulted in a number of recommendations.

38. The Conference agreed that a VMG would normally be comprised of individuals who are knowledgeable with respect to the test method(s) being evaluated in the validation study. Additional expert consultation may be considered periodically during the conduct of a validation study. Depending on the composition of the VMG and the interest its members may have in the method, it may be preferable to engage external individuals who may have a more objective view to the study as consultants. Peer review should be a totally separate scientific process that should be conducted after the completion of the study by a different body whose members are largely independent of the study (examples are those conducted by ECVAM in Europe and ICCVAM in the USA

39. There were extensive discussions regarding the role and responsibilities of the validation management group (VMG). The need for, and role of, a VMG will depend on the purpose of the test and the scope of the validation being undertaken. Therefore, the Conference agreed that a formal VMG structure should not be defined. Instead, the possible roles and responsibilities of the VMG should be clearly defined and described. The primary role of the VMG is to plan the study, organise the logistics of the study, and manage the conduct of the study. Although not all routine activities of the VMG have to be recorded the process should be fully transparent and justifications and rationales for the various decisions of the VMG should be recorded and available, if needed.

40. Although it is normally designated as a group, there are circumstances where the roles and responsibilities of the VMG can be filled by a single individualThe majority of discussions, however, addressed the VMG as a group, because this is seen as the most common manifestation of this role. In the case of a single individual constituting the VMG clear justifications should be presented.

41. There were discussions as to whether, or to what extent, the VMG should be independent of the test(s) considered for validation. There was agreement that the VMG, or its members, do not

necessarily have to be totally independent or disinterested. By virtue of its responsibilities, the VMG will have an interest in the outcome of the validation study, and therefore will carry some degree of bias. Also, the developers or sponsors of a test could be allowed the freedom to serve on the VMG that is designing and supervising the validation of their test.

42. There was general agreement that, although bio-statistics support is necessary for the design and interpretation of validation studies, it was not necessary that a bio-statistician be a formal member of the VMG. The draft Guidance Document is too prescriptive on this point. The VMG, in consultation with statisticians should determine performance criteria of the test that would define the usefulness of the test with respect to its proposed regulatory use. Sources of bio-statistics support could include an external consultant or the lead laboratory statistician. There was agreement that an appropriately qualified statistician be involved in the study design and the selection and implementation of data quality criteria and statistical analytical procedures. Opinion was also expressed that there should be an independent statistical analysis of the overall data.

43. One of the responsibilities of the VMG is the selection and details of the protocol to be used in the experimental work of all phases of the validation. This includes chemical selection, dose selection (where appropriate), endpoints and parameters to be measured, analytical procedures to be used, and the data interpretation procedure.

Recommendation 8:
The many roles of the VMG should be more clearly defined in the Guidance Document. This should include:

- the composition of the VMG, in particular the flexibility with respect to the (in)dependant nature of the membership;
- the involvement of expert consultants including a statistician and
- the monitoring of participating test laboratories.

Recommendation 9:
The Guidance Document should provide examples of how VMG could function; this could also include examples of mistakes made.
Because of the many manifestations of the VMG, “validation management group” should not be capitalised in the Guidance Document.

Expert Consultation by VMG

44. Expert, or peer, consultation was not extensively discussed except to distinguish it from peer review. Expert/peer consultation was considered synonymous with expert/peer involvement. It was emphasised that expert/peer consultation cannot substitute for peer review. It can, however, be used to supply needed expertise in the design, management, and interpretation of the validation study prior to its submission for peer review.

Recommendation 10:
The concept of expert/peer consultation should be addressed in the Guidance Document, and examples should be provided.

Peer Review

45. The Conference agreed that the purpose of the peer review is to determine whether the validation study was properly designed and performed, and that the conclusions of the study are supported by the data generated. The peer review of a method can be carried out in many ways; there is no recommended format for the process.

46. An independent review of the validation study is required prior to submission of a test method for regulatory acceptance or the adoption of a Test Guideline. It was agreed that independent peer review of the validation work is mandatory. This implies that the peer review panel should be independent of the VMG because of the potential bias of the VMG toward the test method and/or the validation study.

47. The Conference further agreed that independence and lack of bias does not necessarily mean that every member of the review panel must be completely independent of the entire test development and validation process, and has absolutely no interest in the topic or test. However, any interest and biases that may be present should be revealed and the panel, as a whole, should be balanced and independent..

48. In addressing paragraph 95 of the Draft Guidance Document and the details of required expertise, the Conference agreed that the necessary expertise needs to reside within the review panel, as a whole, and not necessarily within each individual member. The overall panel should have expertise in the appropriate disciplines for the method under review. These areas of expertise can include, and need not be limited to, validation, technical aspects of the method, statistics, clinical science, general toxicology, etc. It is possible to consult individuals with conflicts to provide necessary expertise to the panel, but these individuals should not be considered as members of the panel and, consequently should not be participating in any discussions related to the overall assessment of the validation study. The process should be unbiased and fully transparent, and the panel must have credibility as an unbiased, knowledgeable group.

49. The meeting was of the opinion that government representatives and regulatory scientist are appropriate peer reviewers provided they have the required expertise, and that their Ministry/Agency does not have a particular interest in the test concerned. The same principle holds for members of any other group who are considered for the panel. Taking into account the requirement for specific expertise how the tests that have been validated, it was agreed that, normally, it would be inappropriate for any standing committee to function as a peer review panel.

50. The meeting agreed that, ideally, the organisation or entity charged with the selection of the peer review and the organisation and management of the peer review should be recognised as independent of the test undergoing validation and scientifically unbiased. Examples of such entities include ICCVAM, ECVAM, national science academies etc. It was recognised that certain organisations have already procedures in place to select peer reviews and manage the process. These procedures may differ between organisations. ICCVAM and ECVAM are composed of government employees who are appointed as representatives of their respective agencies. Similarly, OECD bodies established to independently assess validation studies managed by a VMG, are also composed of appointed government representatives (e.g. the EDTA and WNT).

51. The Meeting was aware that the conclusions of the peer review that a test method is valid does not automatically mean regulatory acceptance. It means only that the test can be recommended to the regulatory agencies for use in regulatory data requirements. A peer review panel may also make recommendations regarding particular uses and potential limitations of the method. The recommendation of the test for regulatory purposes is the responsibility of the test sponsor. It is a separate process that will follow the peer review, and be supported by the peer review report.

52. The Conference discussed extensively whether or not, and to what detail, information from a validation study and subsequent review should be made generally available. In the end it was agreed that preferably all relevant results of the validation study (not necessarily details of participating laboratories or scientists), the peer review process, including the identity and affiliations of the peer reviewers, the peer review report, and the VMG or sponsors response to the peer review report, if any, should be publicly available upon request. Meetings of a peer review panel may or may not be open to the public. All relevant background information and related data should also be available upon request. Where appropriate, the peer review report should be published in a relevant scientific journal.

Recommendation 11:
Any peer review process should, as a whole, be independent of the sponsor and others involved in the validation study, and peer review panel members should not have interests (other than academic) in the test of concern.

Any peer review process should be fully transparent and separate and distinct from the validation process.

Recommendation 12:
Publication of a validation study in a peer-reviewed, scientific journal is highly recommended but cannot substitute an independent peer review process. This should be made clear in the Guidance Document.

Test Development

53. This Guidance Document is not intended to describe the scientific test development process or the process of developing tests as OECD Test Guidelines. Tests come to the validation process by many different paths, including being developed specifically as a regulatory test, or adapted from a test used for other purposes.

54. There was a number of discussions regarding test development. Some of these discussions resulted from the confusion caused by the use of “test development” by OECD in referring to the overall process from unvalidated test method proposal to official adoption of the test as OECD Test Guideline. Some participants were of the opinion that a test that does not have reference data against which its relevance can be measured should be considered “in development” i.e. as being at the stage prior to (pre)validation. It was also noted that initial assessment of the reliability of such a test could already be determined during it’s development. Others were of the opinion that any aspect of validation (relevance and reliability) should follow “test development”. These various opinions illustrated that the process of development, prevalidation and validation should be considered as very flexible.

55. From the discussion it was obvious that there is a "grey" area, by some referred to as "Test optimisation" by others as "prevalidation". Whether or not this part of the process should be considered as part of the validation or test development seems irrelevant. Although there was no full consensus it was also noted that, despite the unavailability of specific reference data in humans or animals, the results of a test that measures a specific biologic mechanism could be extrapolated to humans or animals provided that the mechanism is universal (e.g., basal cytotoxicity; estrogen-receptor binding).

Recommendation 13:
The section of the Guidance Document on test development should be restructured to explain test development issues because much of what is described in this part of the Guidance Document refers to validation and prevalidation.

The Test Sponsor

56. The Conference recognised that every test that is proposed for validation has a sponsor. This sponsor can be the developer of the test, an NGO, or a regulatory authority, or any other organisation interested in having the test accepted for regulatory use. The sponsor is responsible for establishing the validation management committee, providing the logistical support, and for acting on the recommendations of the VMG and the peer review panel.

57. The Conference agreed that the identity of the test sponsor, and the sponsor's role in the validation process, if any, should be disclosed to all who express an interest.

Recommendation 14:
The position and responsibilities of the test sponsor should be described in the Guidance Document.

The Validation Process

58. The Conference recognized that there are various approaches that can be used for test validation, and that the process must be flexible and transparent. Section IV.4.a of the Guidance Document currently provides insufficient guidance on what constitutes appropriate flexibility in the validation process. Participants agreed that it should be made clear that flexibility applies to the scope, the extent, the approach and the organisation of the validation process. However, flexibility should never compromise the degree of scientific rigor needed to properly demonstrate the test's reliability and relevance to the species of interest.

59. The participants confirmed that appropriate numbers of laboratories to participate in the various (pre) validation phases, and the numbers of chemicals to be included in each of the phases cannot be defined in the Guidance Document. These numbers will depend on the specific test being evaluated, the known parameters of the test, and the proposed use(s) of the test.

Recommendation 15:
The Guidance Document should explicitly state that scientific rigor is always required, regardless of the scope of the validation, the type of test, or whether the method is new or revised.

Recommendation 16:
When an *in vivo* study is recommended for validation, statistical support should be used to design an approach that minimises animal use, and to determine the fewest laboratories needed to support the study.

Prevalidation

60. There was general agreement that the concept and process of prevalidation had not been adequately addressed in the draft Guidance Document (see also paragraphs 53-55 on test development). In the draft Guidance Document (paragraphs 38-46), the term "test optimisation" was used interchangeably with "prevalidation". There was general consensus that the concept and process of prevalidation needs to be expanded in the Guidance Document because this phase of the process should be fairly extensive.

Validation

61. It was noted that the definition of the test's performance (mainly it's predictive capacity and to some extent it's interlaboratory reproducibility) should be well characterized prior to the start of a validation study. Normally this characterization is undertaken during the early development and pre-validation stages of a toxicity test's evolution. The formal validation study will most commonly be a multi-laboratory exercise that tests whether or not the new method's predictive capacity and reproducibility meet or exceed success criteria established prior to the start of the study. In this phase chemicals are usually coded. Alternatively a combination of coded and uncoded chemicals could be considered. If results from the prevalidation phase are strong, the formal validation study could be limited with respect to the number of laboratories and chemicals. Irrespective of the approach, the conference agreed that prior to multi-laboratory studies there should be at least preliminary information indicating that the predictive capacity and between laboratory reproducibility of the test is adequate for the stated purpose.

62. Because the purpose of a validation study is the assessment of the relevance and reliability of the test method, there was concern that paragraphs 61-64 of the Guidance Document focus too much on technical details. However, it was also noted that the quality of the laboratories and their performance were critical to the accurate assessment of the test method.

63. The meeting recognised that validation can also be accomplished by computer simulation, as was done for the validation of the up-and-down test (OECD TG 425) by US EPA and subsequently reviewed by ICCVAM. The Guidance Document should explain this option and provide one or more examples.

64. It is possible that data generated subsequent to a test's regulatory acceptance may either change the limitations associated with the test, or call the test's validity into question. For these reasons, the Conference agreed that the validation status of a test might be reconsidered in case new data gives rise to consideration.

Recommendation 17:
The Guidance Document should address more clearly ways to avoid unnecessary testing. These include considering the validation by computer simulation and the decision to start multi-laboratory studies.

Recommendation 18:
The Guidance Document should provide more detailed guidance on how tests can enter the validation process at different stages based on their level of development, prior use, available data, and proposed use.

Validation of Test batteries

65. The Conference addressed the issue of validation of test batteries and testing strategies. The meeting agreed that if a test is to be part of a test battery or testing strategy, the obviously limited biological phenomena it covers should be made clear at the start of the validation process. The Conference confirmed that tests in a test battery should not overlap in order to make-up for weaknesses of other tests in the battery but instead complment each other in order to jointly cover the overall biological phenomenon of interest. The Conference agreed that the individual component tests of the battery or strategy should be validated before considering the validation of the complete test battery.

Recommendation 19:
The Guidance Document should provide more detailed guidance on the roles of tests as components of test batteries and testing strategies.

Validation of QSAR Systems

66. The Conference recognised the need for validation of predictive data models like QSARS. Although time did not allow to go into the specific details that distinguish the validation of QSARS from any other test method, it was recognised that there is a need to provide specific guidance on this issue in the Guidance Document. In particular the recognition that QSARS are not static but progress after each new data entry, needs to be covered. Furthermore, transparency of the pathways and decision steps is of crucial importance.

Recommendation 20:
The Guidance Document should be expanded to provide more guidance regarding the validation of QSARS and other computer-generated systems involving databases than is currently provided in paragraph 90.

"Catch-Up" Validation

67. The term: "catch-up validation" has recently been introduced to describe the validation of tests that are similar to tests that already had successfully undergone validation. The Conference agreed that in these cases the generation of a limited amount of data bridging the new and the validated test may suffice. However, it was also emphasised that in these "catch-up" or "bridging" validation studies, sufficient data should be accumulated to show that the new test performed equally well compared to the previously validated tests.

68. The question was raised as to how a previously not formally validated test that was accepted by regulatory authorities should be validated if a significant test modification was proposed. Although the

issue was not fully resolved, there was consensus that such a modified test would definitely have to be validated, probably through a process similar to "catch-up" or "bridging validation". If the protocol modification was considered substantial, i.e. it would change the data interpretation and the biological phenomenon covered by the test, the new procedure should be considered a new test.

Recommendation 21:
The concept of "catch-up" or "bridging" validation should be explained in more detail in the Guidance Document. This could be illustrated with examples (e.g. the validation of a second *in vitro* corrosion test, very similar to the first one, and the validation of the enhanced guideline 407, adding a number of new endpoints to this existing test).

Retrospective validation

69. The Conference was aware that retrospective validation has been applied to tests that are currently in use (although not for regulatory purposes) by using available data to demonstrate the relevance and reliability of the test. Example of this is the validation study performed by ICCVAM of the local lymph node assay (LLNA) and by OECD of the *in vitro* skin absorption test (TG 428).

70. This was an area of concern and discussion, especially with respect to the retrospective validation of currently used animal tests, which some of the participants noted was not done sufficiently in accordance with currently accepted principles. The current statement in the Guidance Document (paragraph 2) explains that these tests have been accepted by convention based upon their history of use and their demonstrated effectiveness in measuring the respective toxicities.

71. Participants were of the opinion that paragraphs 89 of the Guidance Document as it currently reads may be misinterpreted as if a lower standard could be used for retrospective validations. As discussed in paragraph 58 in addressing flexibility, the meeting agreed that one should not compromise on the scientific rigor of the validation process.

Recommendation 22:
Retrospective validation, i.e., using available data, rather than generating new data, should be described, and explained in a more detail. Examples of successful retrospective validations could be described, e.g., the LLNA and the *in vitro* percutaneous absorption test (TG 428).

Recommendation 23:
The Guidance Document should more clearly indicate than is currently done in paragraph 89 that similar standards apply to the validation of any test, regardless of whether it is a new method with data developed specifically for its validation, or whether the method has been in use and is being validated using available data.

Transparency

72. The issue of transparency was discussed at several sessions and considered by the Conference as a crucial aspect of the process of validation at large. The meeting was aware of the fact that it is difficult to describe the concept of transparency in detail. However, there was agreement that the Guidance Document should cover the issue. Some levels/aspects of transparency are difficult to specify. Transparency, in the context of test validation, requires that all information about the test method and the validation study, including the identity and interest of the sponsor, be available upon request, and that the peer review following the validation be publicly announced. A number of participants were of the opinion that peer review meetings should be open to the public. Others strongly expressed their reservation in this respect. A number of recommendations in this conference report address the issue of transparency; these recommendations appear in the relevant sections of this document.

Recommendation 24:
A section on transparency should be added to the Guidance Document that describes the concept and elements of the process where transparency including public announcement is most relevant.

Reference Chemicals and Data

Reference Chemicals.

73. The participants emphasised that the range of chemicals selected for use in the various (pre) validation phases should be appropriate for the endpoint(s) being measured and the proposed use(s) of the test. As a result, reference chemicals would be selected on a case-by-case basis. The chemicals should be relevant to the event or mechanisms of concern, and to human health or environmental safety. One way of ensuring this would be for the VMG to consult with representatives of the relevant regulatory agencies The experts recommended that the Guidance Document would also address the issue of testing mixtures and whether or not these should be included in the set of reference substances.

74. The Conference agreed that if coded chemicals are used, sufficient information must be provided to allow the chemicals to be tested properly, and so that the safety of the laboratory personnel will not be compromised. Therefore, all necessary physico-chemical and safety information about the test chemicals should be made available to the testing laboratories.
Local and national regulations on the use, storage, and disposal of hazardous chemicals should be addressed. The Guidance Document is rather detailed with respect to coded chemicals and safety for those who handle them.

Recommendation 25:
The GD should include a recommendation to seek input or confirmation regarding chemical selection for validation studies from relevant regulatory authorities.

Reference Data

75. It was agreed that paragraphs 36 and 37 of the Guidance Document need to be expanded to express more clearly that regardless of whether a test is being validated as a new test or as a replacement for a currently used test, its primary purpose is the assessment of human or environmental hazard. Therefore, the reference data should be selected based on a consideration of the endpoint (health effect)

and the species of concern. Participants confirmed that, obviously, the most relevant reference data for new tests are human data. If the goal is to replace an existing animal test the data derived from the "old"test may be the most relevant reference data (as is explained in paragraph 58 of the Guidance Document).However, the Conference emphasised that in case the target species is the human species, priority should be given to good quality human reference data.

76. If new human and/or animal test data are needed as reference data for a validation study, the need to perform the tests should be scientifically and ethically justified. The Conference felt strongly that experimental human studies, other than epidemiological or occupational monitoring studies, should not be performed to develop reference data.

77. It was agreed that where reference data are not available (or insufficient) for the specific endpoint being measured or predicted by the test, the VMG may consider the use of reference chemicals with data derived from related endpoints or to experimentally obtain reference data. Statistical advice should be taken to ensure that the objectives of the study still can be met by using smaller numbers of reference chemicals if the full ranges of chemical types and potencies are not available.

78. There may be newly developed tests for endpoints that were not previously covered by a test. In such cases, there may not be any reference or surrogate data that can be used in the validation study. The Conference agreed that a validation study can indeed be performed on a new test in the absence of appropriate and adequate reference data. The absence of reference data is certainly not a bar to the determination of reliability, whereas the relevance of the test can be at least partially addressed by extrapolation using related reference data. It was proposed that a new test that has undergone such a validation should be termed "provisionally acceptable" to distinguish it from tests that have been validated against appropriate reference data.

79. It was recognised that Member Countries may have different cultures and different ethical considerations, and, consequently, approach ethical questions of testing differently. This will have to be mentioned in the Guidance Document, together with what may be considered to be internationally agreed upon principles.

Recommendation 26:
The Guidance Document should emphasize more strongly that the chemicals selected as reference chemicals should be relevant to the adverse effect or mechanism of concern andalso be relevant for the species of concern.

Recommendation 27:
The need for human reference data should not be used to support human testing, although epidemiologal or occupational monitoring studies would be acceptable. In addition, the development of new animal reference data to validate new test should be strongly discouraged and should only be considered if no alternative is available.

Recommendation 28:
The Guidance Document should address the concept, introduced at the Conference, of "provisionally acceptable" tests for those tests where the relevance could only be partially addressed during the validation because the test's endpoint had not been previously considered.

Animal Welfare

80. Although the Guidance Document does not specifically focus on animal welfare issues, the Conference was of the opinion that the 3Rs should be addressed preferably early in the document.

81. The Conference discussed the preference for validation studies to be reviewed by an animal use committee covering specifically animal welfare considerations. Options are to review the validation study as a whole or each individual component separately. The Conference was aware that not all Member countries may have legislation in place that requires the review of any animal test for animal welfare consideration and, consequently, may not have established animal use committees. It seems preferable, however, that the review of components of the validation studies be done by local committees. A review of the validation study as a whole, if needed, could be conducted by one animal use committee selected by the sponsor of VMG.

Recommendation 30:
The Guidance Document should address up front in the document the importance of fully applying the 3Rs of Replacement, Reduction and Refinement.

Recommendation 31:
The possibility of a review of the validation study or parts thereof by animal use committee(s) should be addressed in the Guidance Document.

GLP Compliance

82. It was noted that there is inconsistency in the draft Guidance Document regarding the need for GLP compliance during validation studies: In paragraph 39 full GLP compliance is highly recommended but not strictly required, whereas paragraph 78 refers to GLP as a requirement. The participants were divided on this issue; some strongly preferred that full compliance with GLP should be required and others were willing to accept compliance with the general principles of GLP, not necessarily all administrative details.

83. The meeting acknowledged that GLP were originally designed for laboratories that are developing data for submission to regulatory authorities. GLP regulations are concerned primarily with the quality of the study conduct including, among other aspects, laboratory record keeping procedures, internal quality control procedures and level of expertise of staff involved. GLP is not with the quality or sufficiency of the underlying science of the test. It was also noted that the first set of GLP were covering *in vivo* studies, and that full GLP compliance includes a number of analyses of test chemicals and test chemical dilutions at various stages of the test. For an *in vitro* test with many chemicals, this could lead to the chemistry analysis being more complex and expensive than the actual conduct of the test.

84. It was noted that the currently used term "in the spirit of GLP" was not clear, and that more guidance was needed. There was general agreement that the most important component of GLP as it pertains to validation studies, was the quality of the conduct of the test and of data records and record keeping procedures. Prior to beginning a validation study, the VMG should make a determination regarding the level of GLP compliance required of the participating laboratories. However, the aim should be full compliance.

Recommendation 32:
The Guidance Document should note that full GLP compliance is preferred, but not always strictly required. However, the Guidance Document should clearly identify the specific procedures that must be in compliance with GLP.

Patented Methods

85. For validation purposes, the Conference adopted the principle that patented methods should be treated like other methods: they should be scientifically valid in order to meet the criteria for regulatory acceptance. This implies that patented tests that function as a black box with our input and output without full understanding of the details of the test, would not be acceptable. The OECD presently does not develop Test Guidelines that require the use of a unique instrument or process owned by a patent. One reason for this is that the method must be readily available to all potential users; another is to avoid market monopoly of an OECD test method by a private company.

86. The performance criteria and other details of a patented method can be generally described, and reference chemicals provided, so that generic methods can be developed and validated. However, this approach might infringe on the patent, and might therefore be illegal. It was noted that because of the difficulty in developing a generic version of the test in each laboratory, and then validating that generic method against the reference chemicals, it would be easier, and probably less expensive, for the laboratories to use the patented method. Even though a generic Test Guideline would be written, the validation study of the patented method would also be referenced in the Guideline. In this way, users of the Test Guideline would be aware of the existence of the patented method and would have the option of using it.

Recommendation 33:
The OECD policy regarding the use of patented methods as Test Guidelines should be described in the Guidance Document together with options how to respect these policies without blocking scientific progress.

Regulatory Acceptance

87. The Conference was well aware that validation and regulatory acceptance are two independent processes and that validation does not necessarily or automatically imply regulatory acceptance. It was agreed that this should be made more clear in the Guidance Document. Regulatory acceptance will only be considered after the validity of the method has been properly addressed. Following the peer review, the report and recommendations are transmitted to the VMG and the sponsor. It was emphasised that it is the responsibility of these recipients to respond to, and possibly follow-up, the criticisms and recommendations of the peer review panel.

88. The Meeting further agreed that if the review is favourable, or after eventual criticism has been addressed, it is the sponsor's responsibility to transmit the test protocol and all supporting information, including the peer review report, to the appropriate regulatory agency(ies) for consideration for regulatory use and submission to OECD for development of the method as an international Test Guideline.

89. After a proposal for regulatory acceptance or Test Guideline development is submitted, the relevant agency or regulatory authority may:

- dismiss the request for regulatory use of the method for reasons other than its validity;
- initiate further review;
- decide that further review is not necessary, or conduct a focussed review;
- start action leading towards international acceptance of the procedure in OECD.

The regulatory authority should provide the reason for any of these decisions.

90. The Conference confirmed that although Member countries and regulatory authorities may have different requirements for regulatory acceptance and may have different ways of converting validated, peer-reviewed tests into regulatory guidelines, all adhere to the criteria for accepting a proposed test for incorporation inanOECD Test Guideline. The meeting emphasised that regulatory authorities should be consulted in the planning of a validation study. This will facilitate the eventual acceptance of the test. The need to consult with regulatory authorities as addressed in paragraph 26 of the Guidance Document may need some additional emphasis.

91. As the Guidance Document is primarily directed towards the incorporation of validated methods in OECD Test Guidelines, it should either include discussion of the mechanism whereby a new OECD Test Guideline is developed and submitted for approval and acceptance or should refer to documents where this is clearly explained. It was recognised that OECD Test Guidelines allow more flexibility, or include options that are usually not considered in detail during the validation of the original test method. This issue, and in particular, the transition from a specific, validated protocol to a more flexible OECD Test Guideline should be clearly addressed in the Guidance Document. The various options for OECD Test Guideline development should all be discussed in the GD. These include:

1) developed/prevalidated ,validated and peer reviewed in OECD.
2) developed and prevalidated outside OECD, but validated through OECD.
3) brought to OECD after validation and converted to a guideline for peer review.
4) brought to OECD after peer review for regulatory acceptance.

Recommendation 34:
More clarity is needed on the link between the peer review of the validation study and the regulatory acceptance process. Although the process of incorporating a newly validated test into a regulatory guideline is outside the scope of the Guidance Document, the link to this process should be addressed.

Recommendation 35:
More practical guidance for submission of a validation study for regulatory acceptance should be added to the Guidance Document to aid in the design of validation studies, possibly in a new section to be inserted after Section VI. As an example of such guidance, it was recommended that the ICCVAM Submission document be appended to the Guidance Document.

Recommendation 36:
There should be either an Annex to the Guidance Document or proper reference to a document that outlines the relevant OECD processes for developing validated test methods as Test Guidelines and Guidance Documents. This should include the role of the National Co-ordinators.

Financial Sponsorship or Support of Validation Studies

91. This topic was mentioned throughout the conference as needing to be addressed. It was noted that it is often difficult to obtain the needed financial support for conducting validation studies. The provision of financial resources for validation studies has not kept pace with the needs and this problem needs to be addressed somehow. The expensive nature of validation studies, as addressed in Joint Meeting document ENV/JM/TG(2001)5 also needs to be recognised in the Guidance Document. The Conference considered it also useful if suggestions could be made for encouraging the sharing of costs, and for active collaboration to avoid duplication of effort. The Guidance Document should be explicit about budgetary issues arising during a validation study and the need for transparency in the provision of funding.

92. It was the general consensus of the participants that this was not an area that could be easily addressed by the Guidance Document, because of the different sources of tests for validation, the rationales for the validation, and differing national and international procedures. The Secretariat suggested that whereas the Guidance Document could not be very specific in this respect, the cover note that would accompany its submission to the Joint Meeting could be more detailed.

Recommendation 37:
The Guidance Document should address the issue of costs involved in validation studies and provide guidance on the most efficient way of using the limited resources.

GENERAL EDITORIAL AND FORMATTING ISSUES

Figures and Tables

93. There were conflicting comments regarding the figures in the draft Guidance Document. It was preferred by a number of participants that the flow charts and terminology used in the Solna document should be used in the Guidance Document in place of the figures currently there, but that less detail and more continuity in the process should be shown. The flow charts should show the sequence of events and decision points from test development to completion of the validation study, and include the stages of peer consultation and peer review. The flow charts should show that the process is a continuous process rather than rigid “blocks” with detailed requirements for entry and exit.

94. The Guidance Document should include references to a variety of case studies and other background information. The flow charts should identify possible entry points into the process for tests in different stages of development, and for different uses. It was emphasized that flexibility is not synonymous with ambiguity, and that flexibility applies to the approach to the validation process and the establishment of reliability and relevance, and not to the principles or the underlying science.

95. An alternative flow chart (see Figure 1 in the Report of Breakout Group 1, included in Annex 11 to this report) was presented that outlines the test development (prevalidation, validation) and the peer review process. There were several disagreements with the ordering of processes in the chart, e.g., whether the determination of reliability should appear before, after, or along side the determination of relevance, and some of the terminology. Although this new chart is a different depiction of the process that is outlined in Figure 1 of the draft Guidance Document, there was no final agreement whether one or both charts should be included. A number of modifications of the flow chart were recommended to its developers. The Conference concluded that those involved in the revision of the Guidance Document should take into account this chart as well as other options.

Definitions and Glossary

96. It appeared that there are inconsistencies throughout the Guidance Document in the definitions of terms. In particular, attention should be paid to consistency of terminology related to the following issues:

- The definitions of test and test method;
- The definition of a test should be broadened to go beyond mechanistic effects
- The terms "new", "revised", "updated" tests should be defined and distinguished;
- The definition of the data interpretation procedure (prediction model) was not accurate. This definition should be revised to reflect the phrase recommended by the meeting,
- The distinction, if any, between replacement and substitution tests should be clarified.

Recommendation 38:
The definitions used should be consistent and where possible, harmonised with those in other internationally quoted documents.

Any additional definitions or proposals for emendations of existing definitions should be forwarded to the Secretariat for inclusion in the Guidance Document.

Editorial and Other Specific Recommendations for the Draft Guidance Document

97. The following suggestions for specific changes were made by Breakout Groups or individual participants. They were not (all) discussed in plenary and should not be considered as agreed by the Conference. However, they should be taken into account when revising the Guidance Document.

- Paragraph 2: There was some disagreement regarding the phrase "…validated based on their history of use…." This section should be changed to reflect the issues addressed in paragraph 69-72 of this report.
- Paragraph 12: The first sentence should be revised to read: "… to obtain information on the adverse effects …."
- Paragraph 13: The definition of a test is also addressed in paragraph 48. The two sections should be harmonised.
- Paragraph 14: This section should be revised somewhat because it addresses only mechanistic tests. Other types of tests could also relevant. A test is described by more than its mechanistic characteristics. There should be discussion and descriptions of apical tests, mechanistic tests, and empirical tests.
- Paragraph 22: The first sentence is very awkward and needs to be rewritten.
- Paragraph 26: Add comments as to why acceptance is based on validation.

- A bridging section is needed between Test Validation (chapter IV) and the Validation Management Group (paragraph IV.1a) to indicate that the overall concepts of the validation process have wide applicability and can be carried out on different scales, from individual to multi-national. Although this document emphasises large-scale validation studies, validation can also be carried out by an individual or small group.
- Paragraph 28 - 32 needs to be revised in accordance with the discussions regarding the VMG composition and roles.
- Paragraph 48 (i): This paragraph is not totally clear and should be revised.
- Paragraph 51: Should be moved to follow after paragraph 53.
- Paragraph 60: This paragraph should take into account the possibility that reference chemicals with a full range of potencies and activities may not be available.
- Chapter IV.4.a. This section does not provide sufficient guidance on what constitutes appropriate flexibility.
- Paragraph 83: This has to be extended to indicate that there are no situations where a lower level of assurance is warranted. What may be needed is less data, or a modified validation procedure.
- Paragraph 89: This has to be revised. The concepts expressed here fall under the headings of retrospective validation or catch-up validation, and should be addressed in their respective sections. Lines 7-8 should be replaced with "For such a case, the assembled data should be evaluated according to the validation principles described above."
- Paragraph 89. The last sentence implies that peer review is optional, whereas at least some kind of peer review was considered mandatory.
- Paragraph 92, footnote. The discussion of peer involvement should be the subject of a separate section and not intermingled with peer review.
- Chapter V.2. This should be titled "Composition of Peer Review Panel."
- Paragraphs 92, 95, and 96 should be combined. The information in paragraph 96 should come before paragraph 95.
- Table 2. Revise title to "Principles and Criteria for Regulatory Acceptance of a New Method."
- Table 2, point a. "… a transparent peer review process."
- Table 2. Add paragraphs: h) Detailed protocols and SOPs should be available; and i) The strengths and limitations of the test should be described.

Other Areas of Concern or Comment

98. The following questions were admitted outside the realm of this Conference, and not addressed by the participants in plenary. They were, nonetheless, discussed among the breakout groups but no conclusions were reached.

- How can the OECD modify its procedures so that new and updated tests can be accepted more rapidly, and would it be possible to adopt Test Guideline if there is not unanimity among Member Countries?
- How can wide-spread and timely acceptance and implementation of new methods be achieved in OECD countries and other countries?

GENERAL CONCLUSIONS AND RECOMMENDATIONS

99. The Conference participants agreed that the Guidance Document provides a good basis for an international consensus on validation and peer review processes. Suggestions and recommendations made by the Conference should be taken into account and the revised draft Guidance Document should be widely circulated for additional comments and suggestions.

100. The Conference considered the validation of new tests for which there are no reference data to compare as one of the most difficult areas to provide guidance. It was agreed that human data should be considered as the "gold standard" but it was also recognised that there are many issues with respect to human data that need further discussion. These include:

(i) validation by retrospective data comparison;
(ii) the ethical issue of human data development;
(iii) the use of occupational monitoring data as an ethical source of comparative human data
(iv) the large number of potential con-founders;
(v) the potential lack of detailed information on chemical exposures.

Recommendation 39:
The Conference strongly recommended that a Workshop or Expert Meeting should be arranged to specifically discuss the acquisition and use of human data as reference data in validation studies.

101. Although the Conference had agreed that the term "Prediction Model" was confusing and subsequently had adopted the description "Data Interpretation Procedure", it also agreed there was insufficient time to provide the background information needed to explain the history of this concept's development. Nor was it possible to fully explain benefits that a clearly stated prediction model provides to toxicity test users and to those conducting validation studies. It was therefore agreed that it would be useful to discuss this very important concept in detail in a separate meeting.

Recommendation 40:
The Conference strongly recommended that a Workshop or Expert Meeting should be arranged to specifically discuss the concept of the "Data Interpretation Procedure".

102. The Conference agreed that the meeting had been very successful: experts from a variety of backgrounds had managed to reach consensus on a large number of issues. The Secretariat explained that the follow-up procedure to the Conference would be as follows:

- The Secretariat with assistance of a consultant (Dr. Errol Zeiger) will provide a draft Report of the Conference to all participants, most probably in May/June 2002.
- Participants will be requested to provide their comments to the draft report taking into account that the report is not intended to be a all-inclusive summary of the discussions but rather a summary of all issues that were agreed.
- Following this commenting round, the report will be finalised and submitted to the WNT for its consideration.

- After acceptance of the report a Drafting Group with two representations from i) EU Regulatory Community, ii) US Regulatory Community, and one representative from iii) Industry, and iv) ICAPO will start the revision of the Guidance Document.
- The revised Guidance Document will again be widely circulated for review before its final adoption by the WNT and its subsequent endorsement by the Joint Meeting.

ANNEX 1

AGENDA

Wednesday, 6 March 2002:

08:30 Conference Pre-meeting: Instructions to Co-Chairs/Rapporteurs

PLENARY SESSION 1: OPENING OF THE CONFERENCE

09:00 Welcome and Opening Remarks.
Ethel Forsberg, Director General National Chemicals Inspectorate KEMI and
Nils-Gunnar Lindquist, Conference Chair

09:10 Historical Perspective, Conference Objectives, Role of Participants.
Herman B.W.M.Koëter, OECD Secretariat

PLENARY SESSION 2: INTRODUCTIONS, SETTING THE SCENE

09:40 Japan's perspectives on the Conference purpose and objectives.
Tohru Inoue, Director, Biological Safety Research Center, NIHS, Japan

10:00 USA and Canada's perspectives on the Conference purpose and objectives.
Susan Hazen, Principal Deputy Assistant Administrator EPA's Office of Prevention, Pesticides and Toxic Substances

10:30 COFFEE/TEA BREAK

11:00 Challenges and Issues Relevant to ECVAM in Relation to the Conference Purpose and Objectives.
Michael Balls, Head of ECVAM

11:20 Challenges and Issues Relevant to ICCVAM in Relation to the Conference Purpose and Objectives.
William Stokes, Director, NICEATM (NTP)

11:40 Challenges and Issues Relevant to Industry in Relation to the Conference Purpose and Objectives.
Mark Chamberlain, Risk Analysis Group Safety and Environmental Assurance Centre Unilever Colworth. Sharnbrook Bedford, UK

12:00 Comments and questions from the audience on any of the presentations

12:20 Instructions to Break-out Groups.
Conference Chair (Nils-Gunnar Lindquist)

12:30 LUNCH BREAK

14:00 **SESSION 1 OF THE BREAK-OUT GROUPS: DEFINING ISSUES, INITIAL DISCUSSIONS**

Break-out Group 1	**Break-out Group 2**	**Break-out Group 3**	**Break-out Group 4**
Principles and criteria for new and updated tests.	Practical guidance on the management and conduct of the validation process	Principles and criteria for regulatory acceptance of validated test methods, including the submission of information to support their validity.	Practical guidance on the process for independent peer review and regulatory consideration and implementation

PLENARY SESSION 3: DISCUSSION OF PROGRESS, DIRECTION AND CROSS-BREEDING OF THE BREAK-OUT GROUPS

17:00 Report from Break-out Group 1 by Rapporteur(s) of Group 1

17:15 Report from Break-out Group 2 by Rapporteur(s) of Group 2

17:30 Report from Break-out Group 3 by Rapporteur(s) of Group 3

17:45 Report from Break-out Group 4 by Rapporteur(s) of Group 4

18:00 ADJOURN FOR THE DAY

Thursday, 7 March 2002:

08:30 **SESSION 2 OF THE BREAK-OUT GROUPS: IN-DEPTH DISCUSSION OF THE ISSUES**

Break-out Group 1	Break-out Group 2	Break-out Group 3	Break-out Group 4

12.30 LUNCH BREAK

14:00 **SESSION 2 OF THE BREAK-OUT GROUPS**

Break-out Group 1	Break-out Group 2	Break-out Group 3	Break-out Group 4

16:00 **PLENARY SESSION 4: DISCUSSION OF DRAFT BREAK-OUT GROUP REPORTS: DISCREPANCIES, CONFLICTING ISSUES,**

16:00 Report from Break-out Group 1 by Rapporteur(s) of Group 1

16:30 Report from Break-out Group 2 by Rapporteur(s) of Group 2

17:00 Report from Break-out Group 3 by Rapporteur(s) of Group 3

17:30 Report from Break-out Group 4 by Rapporteur(s) of Group 4

18:00 ADJOURN FOR THE DAY

CONFERENCE BANQUET (offered by the *Swedish Ministry of Agriculture, Food and Fisheries: Speech by the Mr. Per Goran Öjeheim, State Secretary , Ministry of Agriculture, Food and Fisheries).*

Friday, 8 March 2002:

8:30 **SESSION 3 OF THE BREAK-OUT GROUPS: FINISHING THE BREAK-OUT GROUP REPORTS:**

Break-out Group 1	Break-out Group 2	Break-out Group 3	Break-out Group 4

10:30 COFFEE TEA/BREAK

11:00 **PLENARY SESSION 5: PRESENTATION AND DISCUSSION OF FINAL REPORTS OF EACH OF THE BREAK-OUT GROUPS**

12.30 LUNCH BREAK

14:00 **PLENARY SESSION 5: PRESENTATION AND DISCUSSION OF FINAL REPORTS OF EACH OF THE BREAK-OUT GROUPS** (continued)

15:30 **PLENARY SESSION 6: BRINGING TOGETHER THE REPORTS OF THE BREAK-OUT GROUPS, CONFERENCE CONCLUSIONS, RECOMMENDATIONS AND FOLLOW-UP**

- Bringing together the reports of the Break-out Groups;
- Discussing insufficiently covered or unclear issues;
- Reaching agreement on all major issues;
- Have the objectives been met?
- Conference Conclusions and Recommendations
- Conference Follow-up: OECD process issues

17:00 CLOSING OF THE CONFERENCE

ANNEX 2

OECD Conference on Validation and Regulatory Acceptance of New and Updated Methods in Hazard Assessment

Conférence sur la validation et l'acceptation réglementaire des méthodes nouvelles et actualisées pour l'évaluation des dangers

6-8 March 2002

List of Participants/Liste des Participants
AVAILABLE TO GOVERNMENT REPRESENTATIVES ONLY

ANNEX 3

BREAKOUT GROUP ASSIGNMENTS

1	2	3	4
Breakout Group Chairs			
Phil Botham	Julia Fentem	Otto Meyer	Kathy Stitzel
Dave Hattan	Willie Owens	Len Schechtman	Andreas Gies
Breakout Group Rapporteurs			
Rodger Curren	Rochelle Tyl	Abby Jacobs	Vera Rogiers
Gilly Griffin	Bob Combes	Manfred Liebsch	Karen Hamernik
Breakout Group members			
Leon Bruner	Kailish Gupta	Christiane Aveline	Michael Balls
Ibrahim Chahoud	Ronald Joiner	Odile de Silva	Toini Berzins
Mark Chamberlain	Marike Kolossa	Walter Diembeck	Brita Hagström
Ih Chu	Don MacGregor	M. Dunier-Thomann	Wallace Hayes
Susan Hazen	Tony Maciorowski	Peter Evans	Lena Odland
Taisen Iguchi	Roger McClellan	Jim Freeman	Yasuo Ohno
Mark Jaber	Eva Sandberg	Karin Gabrielsen	Wolfgang Pape
Karl Jensen	J. Riego Sintes	Betty Hakkert	Jennifer Seed
Jun Kanno	Martin Stephens	Ole Ladefoged	Teiji Takei
Dincer Karavut	Tommy Stagh	Gill Langley	Lars Terenius
B. Özturk Kyraci	Sylvie Tissot	Bob Liteplo	Lorraine Twedok
Jean Roch Meunier	Erik Walum	John McArdle	David Wilkins
Ursula Sauer		Edmund Plattner	
Troy Seidle		Tim Springer	
Jan van der Valk		Atsuya Takagi	
Eric Vindimian		Anna Tompa	
Lars Wårngärd		Vanessa Vu	
Neil Wilcox			
Calvin Willhite			
Steering Committee Members and Other Observers			
David Blakey	Bernward Garthoff	David Blakey	Bernward Garthoff
Bernward Garthoff	Alan Goldberg	Bernward Garthoff	Alan Goldberg
Alan Goldberg	Nils Gunner-Lindquist	Alan Goldberg	Nils Gunner-Lindquist
Nils Gunner-Lindquist	Tohru Inoue	Nils Gunner-Lindquist	Bill Stokes
Hiroshi Ono	Horst Spielmann	Gary Timm	Gary Timm
Bill Stokes	Gary Timm		
Andrew Worth	Andrew Worth		
Andrew Worth			

ANNEX 4

BREAKOUT GROUP QUESTIONS

BREAKOUT GROUP 1 - PRINCIPLES AND CRITERIA FOR NEW AND UPDATED TEST METHODS

Co-Chairs: Dr Phil Botham (Zeneca CTL, UK); Dr Dave Hattan (FDA, US)
Co-Rapporteurs: Dr Rodger Curren (IIVS, US); Dr Gilly Griffin (CCAC, Canada)

1. Validation criteria

Background: Validation criteria developed at the 1996 Solna Workshop were based on criteria developed by others (ICCVAM, ECVAM, etc.) and are included in the draft OECD Guidance Document. These validation criteria are to be addressed when determining the usefulness of a proposed test method for a specific purpose.

a. Should and does the Guidance Document provide practical guidance on how test developers should address each of the established validation criteria for each of the situations listed below separately or would more general guidance, or guidance for a few distinct situations, suffice *[see Guidance Document, Table 1, paragraphs 11-16, 18-27, 81-88, 106*]:

- New methods that are proposed to partially replace or totally replace an existing test method?
- New methods that generate safety or hazard data for which there was no prior test method accepted for hazard assessment purposes?
- New methods proposed for use in a tiered testing strategy?
- New methods proposed as a component of a test battery, or multiple test methods proposed as a test battery?
- New methods proposed to provide mechanistic information?
- Revisions of existing methods, including the addition of a new endpoint(s) to an accepted method?
- Should the criteria be the same for screens *(in vitro, in viv*o) and replacement tests *(in vitro, in viv*o)?
- Promising methods in use but without sponsors and / or data collected with more than one protocol.

b. Where there is not considered to be adequate guidance, what additional guidance would be helpful (for each of the validation criteria)?

c. Is it possible to indicate to what extent an existing protocol can be modified before a (new) validation study would be required to assess the scientific validity (relevance and reliability) of the modified protocol*?[see Guidance Document, paragraphs 84-85, 106]*.

d. Does the validation of a test guideline that is more generic than a test protocol require other or additional considerations to ensure that the guideline is as relevant and reliable as a specific test protocol? *[see Guidance Document, paragraphs 11, 21, 81-83]*.

2. Assessing the test method validity: test variability and validation

Background: Historically (animal) tests were developed for endpoints of concern in humans by studying these same endpoints in animals (e.g., acute toxicity seen at a particular dose level in the rat was considered as also acutely toxic in humans). "Relevance" was not an issue, "reliability" was often not even considered. The validation of tests that require several extrapolation steps (e.g., *in vitro* eye irritation) has strongly focused on prediction models and the inclusion of "gold standards" to proof the "relevance" of the test. In addition, the "reliability" of these tests is usually addressed in great detail.

Should there be a difference in validation approaches (with respect to the relevance of the test) depending on the purpose of the test, its level of standardisation and the type of observations/measurements relative to the subject of interest? *[see Guidance Document, paragraphs 8, 12-16, 33, 34,35, 39, 50-53, 56-60, 70-74, 81-83, 89, Tables 1 and 2]*.

Reliability

a. Should the level of scrutiny with respect to the assessment of the reliability of a test depend on the complexity of the extrapolation of the test results*?[see Guidance Document, paragraphs 8, 12-16, 81-83]*.

b. Can guidance be provided as to the range of number of chemicals and laboratories that are needed for determination of intralaboratory variability? *[see Guidance Document, paragraphs 39, 50-53, 56-60]*.

c. Can guidance be provided as to the range of number of chemicals and laboratories that are needed for determination of interlaboratory variability? *[see Guidance Document, paragraphs 39, 50-53, 56-60]*.

d. What statistical approaches should be applied to assess reliability? *[see Guidance Document, paragraphs 33, 34, 70-74]*.

Relevance

a. Should the level of scrutiny with respect to the assessment of the relevance of a test depend on the complexity of the extrapolation of results? *[see Guidance Document, paragraphs 8, 12-16, 81-83]*.

b. How is the appropriate current standard selected to compare the new or alternative method to; what guidance may be provided for a method for which there are no reliable, equivalent, *in vivo* data? *[see Guidance Document paragraphs 14-15, 35, 37, 45, 48, 81-83]*.

c. What guidance can be provided about documenting and interpreting the relationship and meaning of a test observation to the species of interest, including limitations with regard to *[see Guidance Document, paragraphs 12, 14, 35, 108, Table1, 2*]:

- Nature of response;
- Health impact of responses;
- Correspondence of responses;
- Dose level/dose response;
- Timing of response/time action;
- Reversibility of response;
- Species/strain/sex differences.

BREAKOUT GROUP 1 – ALTERNATIVE SET OF QUESTIONS

1. Validation criteria

Background: validation criteria developed at the 1996 Solna Workshop were based on criteria developed by others (ICCVAM, ECVAM , etc.) and are included in the draft OECD Guidance Document. These validation criteria are to be addressed when determining the usefulness of a proposed test method for a specific purpose.

a. Does the Guidance Document provide adequate practical guidance on how test developers should address each of the established validation criteria for each of the situations listed below:

- New methods that are proposed to partially replace or totally replace an existing test method?
- New methods that generate safety or hazard data for which there was no prior accepted test method?
- New methods proposed for use in a tiered testing strategy?
- New methods proposed as a component of a test battery, or multiple test methods proposed as a test battery?
- New methods proposed to provide mechanistic information?
- Revisions of existing methods, including the addition of a new endpoint(s) to an accepted method?
- Should the criteria be the same for screens *(in vitro, in viv*o) and replacement tests *(in vitro, in viv*o)?

b. In each of the above situations where there is not considered to be adequate guidance, what additional guidance would be helpful (for each of the validation criteria)?

c. To what extent can an existing protocol be modified before a new validation study would be required to assess the scientific validity (relevance and reliability) of the modified protocol?

d. In adopting a test guideline on the basis of a validated test protocol, what steps should be followed to ensure that the resulting, more-generic, guideline is valid?

2. Assessing test method validity: standardization vs. validation studies

***Backgroun*d:** historically, *in vivo* toxicological test methods underwent a standardization process to demonstrate their practical utility. The method was tested with a limited number of chemicals of different hazard potential, to ensure that it was suitable. More recently, detailed validation procedures have been articulated for demonstrating the relevance and reliability of methods.

Test standardization

a. Is there merit in continuing test standardization for *in vivo* test methods?
b. If so, which tests (e.g. costly, long-term studies)?
c. Is such standardization adequate in determining the usefulness of at least some *in vivo* methods?
d. Is test standardization adequate for *in vitro* methods?
e. How does test standardization differ from prevalidation, and could prevalidation be sufficient on some occasions (define which) to demonstrate validity ?

Validation

a. Are detailed validation criteria (test relevance and reliability) and processes valuable tools in assessing test method validity?
b. For which tests - *in vivo, in vitr*o, both?
c. How closely should detailed validation approaches be followed?
d. Would significant deviation from these lead to non-acceptance of the proposed test in Member Countries?

General

a. Is there merit in using both test standardization/prevalidation and detailed validation practices?
b. If so, which validation criteria should be considered for each?
c. How can the validity of a method for which there are no/inadequate data available for comparative purposes be assessed? (e.g. novel endpoint, or an *in vitro* test for which there are no reliable, equivalent, *in vivo* data)

3. Test relevance

Background: one part of knowing the usefulness and limitations of a test deals with test relevance. Traditionally, there are two components. One is a scientific determination - what is the meaning of the observations in the test under review to observations in the species of concern? The other is a pragmatic issue, namely what is the relevance of test outcomes to a given regulatory program?

Test observations to species of interest

What guidance can be provided about documenting and interpreting the relationship and meaning of test system observations to the species of interest, including limitations with regard to:

a. Nature of response
b. Health impact of responses
c. Correspondence of responses
d. Dose level
e. Dose response (when applicable)
f. Timing of response
g. Time action (when applicable)
h. Reversibility of response
i. Sex differences
j. Strain differences
k. Species differences

Test observations to regulatory programs

What guidance can be provided about documenting and interpreting the relevance of a test method outcome to a given regulatory program, with regard to:

a. Is the health effect of interest to the program?
b. Is the test measure of importance to the program?
c. Is the test observation adequate to meet program needs?

General

a. What statistical approaches should be applied to assess relevance?
b. What is the role of the prediction model?

4. Test variability and determination of reliability

Background: one important component of a test's validity deals with the reliability of the method. The number of chemicals under test, and the number of laboratories involved, are variables. In some cases, authorities simply have a different mind set as to the design of validation studies; in others, considerations of cost and time are paramount.

Considering: (a) *in vitro* methods, (b) short-term *in vivo* methods and (c) chronic *in vivo* methods:

a. Can guidance be provided as to the range of number of chemicals that are needed for determination of intralaboratory variability?

b. Can guidance be provided as to the range of number of laboratories that are needed for determination of interlaboratory variability?

c. Can guidance be provided as to the range of number of chemicals needed for determination of interlaboratory variability?

d. What statistical approaches should be applied to assess reliability?

BREAKOUT GROUP 2 - PRACTICAL GUIDANCE ON THE MANAGEMENT AND CONDUCT OF THE VALIDATION PROCESS

Co-Chairs: Dr Julia Fentem (Unilever, UK); Dr Willie Owens (Procter & Gamble, US)
Co-Rapporteurs: Dr Bob Combes (FRAME, UK); Dr Rochelle Tyl (RTI, US)

1. Validation programme needs and responsibilities

Does the Guidance Document provide adequate guidance on the different needs and responsibilities in a validation programme, such as:

a. *A priori* agreement on the regulatory need and purpose of the proposed test in order to consider and design an appropriate validation programme? *[see Guidance Document paragraphs 11,110,111*].

b. Possible consultation on the proposed protocol with experts and national authorities? *[see Guidance Document paragraph 26*].

c. The independent management of the validation programme (including composition of a validation management group and other steering bodies)? *[see Guidance Document paragraphs 28-32*].

d. Independent review of the validation programme results? *[see Guidance Document paragraphs 91-101*].

e. The acceptance (or rejection) of the test protocol by national regulatory authorities? *[see Guidance Document paragraphs108-111 and Table 2*].

2. Validation programme design

Does the Guidance Document provide adequate guidance for the design of validation programmes, such as:

a. The overall approach and possibility of dividing the validation programme into phases? For example, consideration whether a test optimisation (or protocol comparison) phase is necessary, or consideration of phases to address increasing level of protocol specificity, or the contrary, greater protocol flexibility? *[see Guidance Document paragraphs 7,24-25, 38, 40-49, 81-90 and Figure 1]*.

b. The different types of tests and the different purposes for a test that may be considered, such as *in vitro* and *in vivo* assays, screens and replacement tests, a component of a battery and the battery itself, and different tiers or steps in a tiered scheme? *[see Guidance Document paragraphs 14-16, 48-49, 81-90]*.

c. When is there a need for comprehensive background documents on a test (including the scientific basis for the method, its history and state of development, perspective on current or alternative tests, protocol variations employed and not employed, and chemicals previously used in the test and their results)? *[see Guidance Document paragraphs 11, 14 20-21 and Figure 1]*.

d. The type and variety of chemicals needed to assess the relevance and reliability of the test? *[see Guidance Document paragraphs 36-37, 56-60]*.

e. How can computer simulations be used in the design of the validation study (e.g. in the assessment of the maximal predictive performance that can be expected of a new test, taking into account the error associated with the data sets and the power of the protocol)?

3. Planning of a validation programme

Does the Guidance Document adequately address the planning of a validation programme?

a. Is there an adequate description of the tasks that should be performed and data provided to the independent reviewing body in order to validate a study (this issue will be extensively addressed by Breakout Group 3)? *[see Guidance Document paragraphs 11, 75-80, 91-104]*.

b. Should the planning of a validation study always be publicly announced? When is public announcement essential and for what reasons? In those cases where public announcement is considered useful, should public comments be solicited ?

c. What should be the involvement of national and/or international regulatory authorities and industry in the planning and conduct of a validation study? Should institutions/centers experienced in conducting validation studies be involved in each validation project? *[see Guidance Document paragraphs 28-32]*.

4. Promotion of validation programmes and sharing the financial burden / who should validate?

In the past, test methods have been developed and their scientific attributes evaluated by different authorities, including Member Countries, OECD, and local and regional authorities. In certain cases, such as the local lymph node assay, the development and validation was conducted by industry and the programme results submitted to ICCVAM for review and then to national authorities and the OECD for acceptance. Validation programmes usually require considerable financial resources and time of the parties conducting the programme. In light of these considerations:

a. Should test development and validation remain flexible and open to Member country authorities and other parties with sufficient expertise, to initiate this work or should test development and validation be limited to a number of recognised authorities, who would become specialised in this work ?

b. Should any test method proposal to the OECD be already validated, or could validation be part of the OECD Test Guideline development process? *[see Guidance Document paragraphs 11, 20]*.

c. Should there be an "International Validation Initiative" analogous to the Human Genome Project to divide the work? If so, how could this be accomplished?

d. Should the OECD continue to undertake, on occasions, validation efforts such as with the uterotrophic and Hershberger assays when requested by Member countries? What is the best way to have budgets that are adequate to meet the desires of Member Countries to develop OECD guidelines? Should government or industry provide funds for validation?

5. Chemicals selection

Validation studies are quite often limited by the availability of chemicals backed by high quality in vivo data both from testing in animals or clinical human data. Are the availability, selection and numbers of reference chemical substances adequately addressed in the Guidance Document? Issues may include:

a. What is the adequate number of positive and negative test chemicals to validate a given protocol? *[see Guidance Document paragraphs 36-37, 56-64]*.

b. What are the possibilities to combine *in vivo* data from animals and humans? *[see Guidance Document paragraphs 36-37]*.

c. The usefulness of establishing chemical repositories and databanks for validation studies, so that adequate quantities of the same lots can be used and the chemicals prepared, coded and distributed?

d. Would it be useful to establish a list of (inter)national experts and advisors for chemical selection and repository procedures to support validation management groups?

6. Information provided to laboratories in a blind trial

Blind (coded chemicals) trials are usually considered necessary in assessing the validity of a test. Is blind testing sufficiently addressed in the Guidance Document? Issues for consideration include: a. In a blind trial, should specific concentrations or pre-determined ranges be employed in the protocol (since this will improve biostatistical analysis and performance of the test), or should only information on solvent and solubility be provided allowing the laboratories to choose their own concentrations or doses? *[see Guidance Document paragraphs 61-64]*.

b. Since occupational safety is an important issue in laboratories during blind testing, how should the safety information be made available in the protocol and, particularly, in possible emergencies, without compromising the study? *[see Guidance Document paragraphs 61-64]*.

7. Selection of laboratories

a. Does the Guidance Document provide sufficient guidance on criteria and other issues related to the selection of qualified laboratories to participate in the validation programme? *[see Guidance Document paragraphs 39, 50-55]*.

b. Issues to be considered include: GLP status, training and technical proficiency of laboratory staff prior to the start of a validation study, geographical representation, and equipment and facilities required to conduct the test. *[see Guidance Document paragraphs 65-66]*.

8. Biostatistical requirements

The sound biostatistical evaluation of the data is central to a validation programme.

a. Should the statistical evaluation of the data continue in the future to be performed by an independent biostatistician as in previous programmes? What are the advantages and disadvantages? *[see Guidance Document paragraphs 30, 33-34]*.

b. Should the biostatistical analysis be in accordance with GLP? *[see Guidance Document paragraph 39]*.

c. Does the Guidance Document provide sufficient guidance on how to select a statistician and how to review and approve the statistical analyses of the validation programme data? *[see Guidance Document paragraph 30]*.

d. Does the Guidance Document provide sufficient guidance on the development of biostatistical prediction models (PMs)? *[see Guidance Document paragraph 35]*.

e. Who should determine the performance criteria for a valid test (including the number of test chemicals) and whether the PM chosen will enable the expected performance to be assessed -the management team or the biostatistician?

f. Should the data collection and submission always be done in an electronically standardized format by the participating laboratories, in order to facilitate the handling and analysis of the data? *[see Guidance Document paragraphs 67-74]*.

BREAKOUT GROUP 3 - PRINCIPLES AND CRITERIA FOR REGULATORY ACCEPTANCE OF VALIDATED TEST METHODS, INCLUDING THE SUBMISSION OF INFORMATION TO SUPPORT THEIR VALIDITY

Co-Chairs: Dr Leonard Schechtman (FDA, US); Otto Meyer (Division of General Toxicology, Denmark)
Co-Rapporteurs: Dr. Abby Jacobs (FDA, US); Dr Manfred Liebsch (ZEBET)

1. Acceptance criteria

Background: Regulatory acceptance criteria were developed at the 1996 Solna Workshop, and were based on criteria developed by others (ICCVAM, 1995; European Commission, etc.).

a. Should and does the Guidance Document provide adequate practical guidance on (1) how test developers should address each of the established acceptance criteria for each of the situations listed below separately, and (2) how regulatory authorities should assess the extent that these criteria have been met *[see Guidance Document paragraphs 4-5, 11-16, 81-89, 103-104, 107-108 and Table 2*]:

- new methods that are proposed to partially replace or totally replace an existing test method?
- new methods that generate safety or hazard data for which there was no prior accepted test method?
- new methods proposed for use in a tiered testing strategy?
- new methods proposed as a test battery?
- new methods proposed to provide mechanistic information?
- revisions of existing methods?
- in vivo methods?
- in vitro methods?
- ecotoxicology methods?

b. Where the guidance is considered to be inadequate, what additional guidance would be helpful?

c. Are the regulatory acceptance criteria as included in the Guidance Document adequate and appropriate for any or all of the types of test methods *(in vitro*, *in vivo*) listed under 1a? Are there additional criteria that should be added for any or all of the types of test methods listed under 1a? *[see Guidance Document paragraphs 103-104, 108-109 and Table 2]*.

d. Are the OECD acceptance criteria essentially similar to other established regulatory acceptance criteria and, if not, how can the OECD acceptance criteria be updated, revised and harmonized? *[see Guidance Document paragraphs 4-5]*.

e. Is independent peer review of the validation study mandatory for regulatory acceptance? If not mandatory does it enhance the chances of regulatory acceptance of a method?

2. Submission guidance

***Backgroun*d:** Regulatory authorities evaluate the extent that proposed test methods address each of the established principles and criteria for validation and regulatory acceptance. Detailed submission guidelines have been developed that provide guidance to test developers on what information should be provided, as well as a systematic format for organizing such information (e.g. ICCVAM, 1999).

a. Guidance on test submission to the regulatory authorities should facilitate their evaluation of new test methods. Are currently available submission guidances considered helpful by regulatory authorities and do they provide helpful guidance for test developers in organizing information and data for submission?

b. Does the Guidance Document provide sufficient guidance with respect to the submission of a test proposal? If not, would appending the ICCVAM and ECVAM outlines for submissions to the Guidance Document further enhance it and provide the desired additional guidance or would (extended) general guidance suffice? Would use of these existing detailed frameworks help alleviate the need to draft separate (duplicative or replacement) guidance for the OECD Guidance Document? *[see Guidance Document paragraphs 11, 75-80, 91-104]*.

c. Should there be a requirement for minimum information for data submissions for new or updated test methods?

d. Would it be useful to harmonise submission guidance at the international level? How might this be achieved?

e. Could test methods be considered for regulatory use irrespective of the level of detail of the documentation explaining the extent to which each of the validation and acceptance criteria have or have not been addressed? *[see Guidance Document paragraphs 103-106]*.

f. With respect to promoting international acceptance of adequately validated test methods, what procedures would be useful to encourage acceptance by regulatory authorities?

3. Relevance and reliability criteria for regulatory acceptance

Are there any or should there be any stringent acceptance criteria for determination of test method relevance and reliability, e.g. number of laboratories, number of test chemicals, inter-and intra-laboratory variability, test method performance? *[see Guidance Document paragraphs 20, 22-27, 33-74 and Table 1]*.

4. Improvement/modification/amendment of data submitted and/or need for additional work

If the data set of a validation study is considered inadequate, additional work may be warranted. Should there be guidance on who should be responsible for the additional work? *[see Guidance Document paragraphs 105-107]*.

5. Patented test methods

***Backgroun*d:** Patented test are becoming more and more common; commercial test kits are available to screen for health hazards; transgenic strains are being developed; gene chip and other new informatic methods are being assembled. Patented tests have been judged to be acceptable locally (U.S. transport, pharmaceuticals, pesticides and industrial chemicals) and regionally (EU). Current OECD policy necessitates that such methods are described in general terms and in a way that the test could also be conducted without the need to obtain the patented materials. *[see Guidance Document paragraph 104 and tables 1b and 2e*].

a. What further guidance could be provided in order to allow the use of patented tests or test materials while avoiding that an OECD Test Guideline becomes dependent on a particular supplier of the patented materials?

b. Will current OECD policies with respect to patented procedures inhibit the development of scientifically promising methods?

6. Test Guidelines vs. Validated test method protocols

Background: OECD Test Guidelines and other national or international test guidelines are flexible, allowing for some case-specific latitude in the conduct of the test. This might include the provision of a selection of choices (e.g. species, stock/strain, gender, tissue source, tissue viability, diet, fasting/non-fasting), allowing ranges (e.g. age, weight, room humidity and temperature), and/or not providing specific technical or procedural detail (e.g. histopathological observations, clinical observations, positive controls, specific records). Adding test model and procedural/technical flexibility in test guidelines could result in data being generated from significantly different protocols than those that have been validated. Deviations from the validated protocol could potentially yield altered predictions of hazard, e.g. under- or over-prediction of toxicity (increased false negatives or false positives, respectively).

a. Does the Guidance Document provide sufficient guidance for the design and conduct of validation studies (to cover the identification and inclusion of essential protocol variables in the validation work) that would allow for the necessary protocol variables in the Test Guideline without deviating significantly from the validated method? *[see Guidance Document paragraphs 23,25,38,47,105-108,110-111]*.

b. Should test guidelines incorporate specific validated protocols? Should users be required to substantiate the validity of any modifications introduced if they choose to modify the validated protocol? If so, to what extent?

c. If regulatory authorities allow for use of protocols that employ modified procedures or materials (i.e. a “general” test guideline) that differ from the validated protocol, should the validity of those modifications be substantiated before incorporating such flexibility in test guidelines?

d. Prior to regulatory acceptance of a general test guideline that is based on a specific validated protocol, what documentation should be provided to support the validity of the flexibility incorporated into the general test guideline?

BREAKOUT GROUP 4 - PRACTICAL GUIDANCE ON THE PROCESS FOR INDEPENDENT PEER REVIEW AND REGULATORY CONSIDERATION AND IMPLEMENTATION

Co-Chairs: Dr Andreas Gies-Reuschel (Umweltbundesamt, Germany); Dr Kathy Stitzel (Procter & Gamble, US)
Co-Rapporteurs: Dr Karen Hamernik (EPA, US); Dr Vera Rogiers (Vrije Universiteit Brussel, Belgium)

1. Independent peer review

Background: Independent peer review processes are general practice for the acceptance of scientific papers and it is often recommended to consider an independent review of validation studies as well. The independent nature of the management of validation studies also appears to be crucial.

Does the Guidance Document sufficiently cover the issue of independent management and review of validation studies? The following questions are in particular relevant *[see Guidance Document, paragraphs 28-32, 75, 91-102*]:

a. Is independent peer review essential for validation studies at all times?

b. In case the validation study is managed by an independent validation management committee, is peer review still necessary? If yes, what would be gained from it?

c. In case a validation study has been performed by, or under the auspices of, a particular center, could the same center be involved in the peer review process, if deemed necessary?

d. Should a peer review be conducted by experts with respect to the test that has been validated or could it be conducted by general experts in the area of validation?

e. Are formally nominated government representatives considered as independent

f. Could a standing committee of nominated government representatives function as peer review committee?

2. Regulatory considerations [see Guidance Document, paragraphs 91-102, 104].

a. In case a regulatory authority requires a specific data set and has expressed confidence in a test method that provides that information, does this waive the need for an independent peer review of the validity of that test method? If yes, should this decision be justified in writing? If not, what could be gained from the peer review?

b. Should representatives from the regulatory authority of interest participate in the peer review process?

c. Obviously, the implementation of a new test as an accepted means to fulfill a specific data requirement is the prerogative of the regulatory authority. Could these authorities be forced to accept the validated method? Even when the new method was validated/reviewed elsewhere and not internationally accepted?

d. When should the update/revision of an available test method be considered for peer review? At all times or only when the changes made to the method are substantial? And what is substantial?

e. Are there special committees established for the acceptance of new methods at the competent regulatory agencies in Europe, Japan and the USA?

f. Are they endorsing methods for general use for regulatory purposes or also for specific use under specific areas of regulation, e.g. only for cosmetics, pesticides, etc.?

g. What are the roles of the ICH and the OECD in this process?

h. How long does it usually take from successful peer review of a validation study to international regulatory acceptance and implementation into national legal practice?

3. Publication

The number of scientific journals which publish the full results/data sets of validation studies is limited. Moreover, reviewers are usually not familiar with the harmonised OECD criteria for conducting validation studies, in particular as far as the biostatistical criteria for assessing the outcome of a validation trial is concerned.

a. What are the appropriate journals for publishing the results of a validation study? *[see Guidance Document, paragraph 102]*.

b. Is there an advantage or disadvantage in publishing every phase of a comprehensive validation study separately (shortly after the results have been obtained and reviewed, e.g. for the prevalidation)?

c. If publication in the scientific literature takes too much time or does not seem appropriate, would review by a panel of experts suffice? *[see Guidance Document, paragraphs 95-97]*.

d. Alternatively, should peer review for regulatory purposes by an expert panel be required if the results of a study have been published in the peer reviewed literature? *[see Guidance Document, paragraphs 95-97]*.

4. Existing processes - peer review or peer involvement?

Background: Two methods have been traditionally used to evaluate the scientific underpinnings of new test methods: peer review is independent evaluation by experts in the area of the test under review who are uninvolved with the test and its validation; peer involvement is evaluation of the method by parties involved with the method (e.g., industry, government, public interest). Both types of review are currently applied and in some countries the independent evaluation may also involve a public advisory committee of people who may or may not be independent and considered as experts.

a. What are the pros and cons (e.g., scientific rigor, policy analysis) for the processes used by different authorities?

b. Do the different processes yield similar decisions as to the adequacy of tests (e.g., standardization, reliability, relevance)?

c. Should a statement of validity of a method by one of the above authorities be considered adequate for all other authorities?

d. Should Member Countries or regional authorities use the same type of review process before sending a method to OECD for consideration?

e. Is there an optimal process that is realistically doable?

5. Transparency

Background: Authorities vary in the degree of openness of proceedings and decision-making. Some are essentially closed processes, open to but a few parties, others are open to public input. The transparency with respect to documentation, including justification of decisions made, also may vary from region to region.

Does the Guidance Document provide sufficient guidance with respect to transparency of the validation and regulatory acceptance processes and decision making? *[see Guidance Document, paragraphs 78-80, 102]*.

6. Communication and mutual acceptance [see Guidance Document, paragraphs 110, 111].

a. Are there any mechanisms in place to inform industry and regulators about changes and progress in replacing regulatory tests with new methods?

b. Are all international regulatory agencies informed about the acceptance of a new method after it has been accepted by one agency, and is there mutual acceptance of a new validated test?

c. Which role can/must the OECD play in this process?

ANNEX 5

Presentation made by: Herman B.W.M. Koëter of the OECD Secretariat

OECD Conference on Validation and Regulatory Acceptance of New and Updated Methods in Hazard Assessment

6th-8th March 2002
Stockholm, Sweden

OECD 1 OCDE

OECD's first priority in hazard assessment:

Human Health and Environmental Safety

OECD 2 OCDE

Purpose of the Conference

"Develop (and achieve consensus on) practical guidance on principles and processes for the validation and acceptance of test methods for regulatory hazard assessment purposes."

OECD 3 OCDE

Conference Objectives:

Provide practical guidance on how to address:

- the established **validation principles and criteria**;
- the conduct and management of the **validation process**;
- the established **regulatory acceptance principles and criteria**; and
- the **process for regulatory implementation** (including independent review).

OECD 4 OCDE

Conference Outcome

Reach consensus on recommendations for the improvement and completion of draft Guidance Document No. 34

OECD 5 OCDE

The Conference is about internationally agreeable guidance on:

1. Producing evidence that a testing method is for hazard assessment are relevant and reliable and ready for use in regulatory assessment (**how to do it**);
2. Providing the evidence to the regulatory authorities in a way that facilitates the acceptance and use of the validated method in regulatory hazard assessment (**how was it done**).

OECD 6 OCDE

Keeping in mind the Conference objectives and aim:

- Let's not re-invent the Solna Principles and Criteria;
- Let's be practical and focus not too much on "principles" only;
- Remain within the framework of regulatory assessment needs;
- Focus on science, rather than policy.

OECD 7 OCDE

The Conference is not focussed on:

- the process of OECD Test Guideline development;
- whether or not OECD Test Guidelines should allow for protocol variables;
- on Good Laboratory Practice rules and guidance;
- on test methods used in fundamental research or "in house" priority setting;
- animal welfare/in vitro assays and other alternative tests.

OECD 8 OCDE

In order for the Guidance Document to be internationally acceptable...

- It cannot be very specific but should also not be too general: find the balance!
- it should cover all essential elements of the validation process;
- it should cover all essential elements needed to satisfy the regulatory community;
- it should comply with current policies in OECD Member countries; and...

OECD 9 OCDE

In order for the Guidance Document to be internationally acceptable...

- → It should provide adequate guidance on flexibility. Issues include:
- full transparency;
- allowing for social and cultural differences;
- harmonisation, not standardisation;
- no two validation studies are the same;
- co-operation; sharing the work is not the same as "bifurcation".

OECD 10 OCDE

Keep in mind that without flexibility we would not have regulatory agreement today on:

- three alternative methods for TG 401;
- in vitro dermal corrosivity tests;
- the Local Lymph Node Assay (LLNA);
- in vitro phototoxicity test;
- and tomorrow on an in vitro percutaneous absorption test.

Other relevant issues (1):

- Experts do not represent their country but participate in their personal capacity;
- Experts are all requested to actively participate in the discussions;
- The Breakout Groups are the heart of the Conference...

Other relevant issues (2):

Breakout Groups are requested to consider:

- please stay focussed on purpose, objectives and aimed outcome of the Conference;
- main references are the draft Guidance Document and the compilation of comments;
- the sets of questions are a tool; to answer them is not the Conference objective;
- use supporting and background documents as intended: background, not discussion.

OECD 13 OCDE

Other relevant issues (3):

- the Conference report will cover:
  - the essence of the plenary lectures (slides will be annexed);
  - summaries of the most relevant plenary discussions;
  - brief reports of the breakout groups including answers to the questions;
  - detailed suggestions for improvement of the draft Guidance Document;
  - other recommendations as appropriate.

OECD 14 OCDE

Other relevant issues (4):

- The Conference Report will be circulated to all participants for approval;
- Some participants will hopefully volunteer to take the lead in the revision of the draft Guidance Document;
- The revised draft Guidance Document will be widely circulated for comment together with the approved report of the Conference.

OECD 15 OCDE

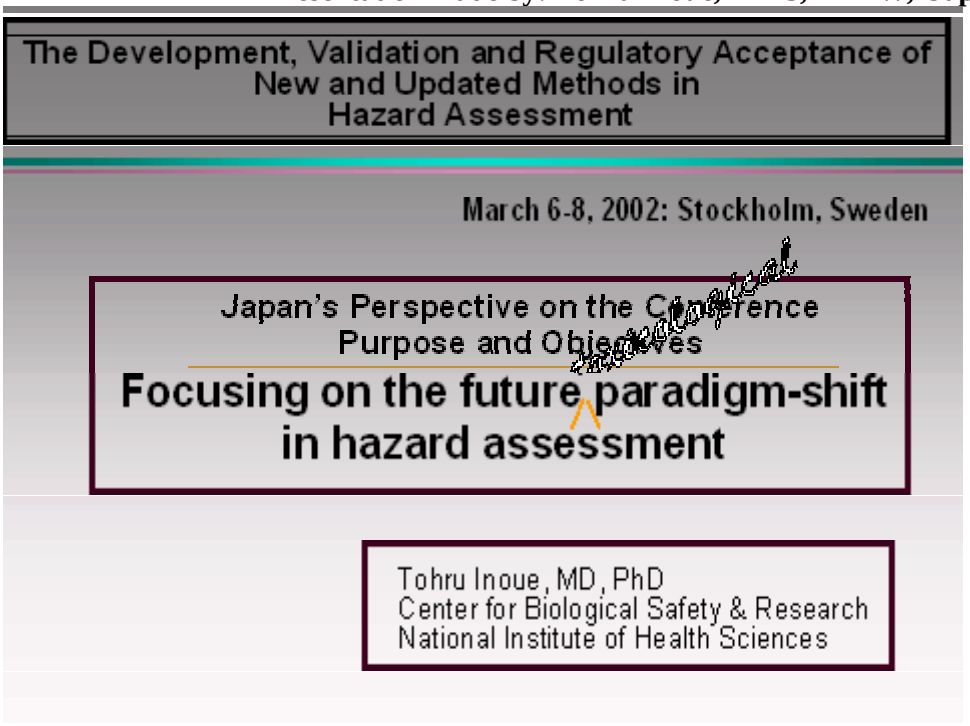
"It is the mark of an educated mind to rest satisfied with the degree of precision which the nature of the subject permits, and not to seek an exactness where only an approximation of the truth is possible"

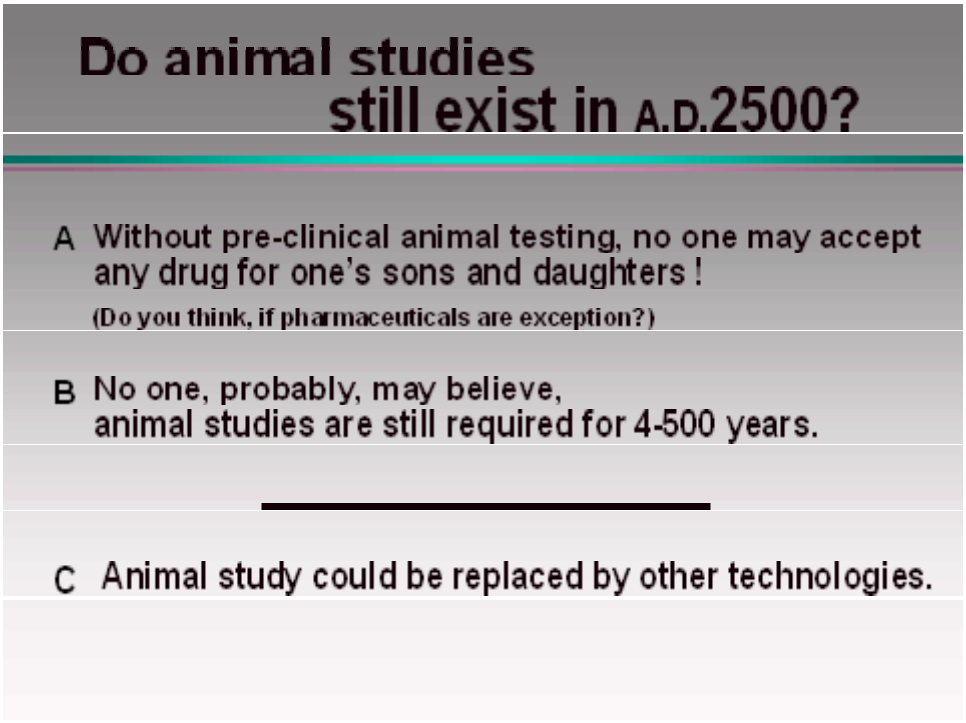
(Aristotle)

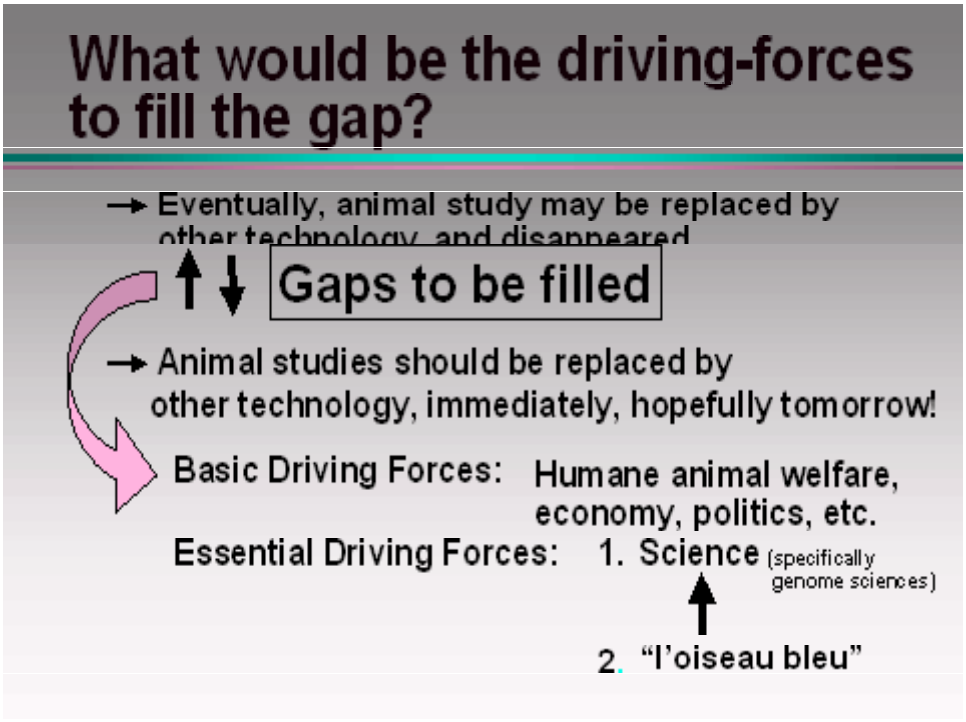
OECD 16 OCDE

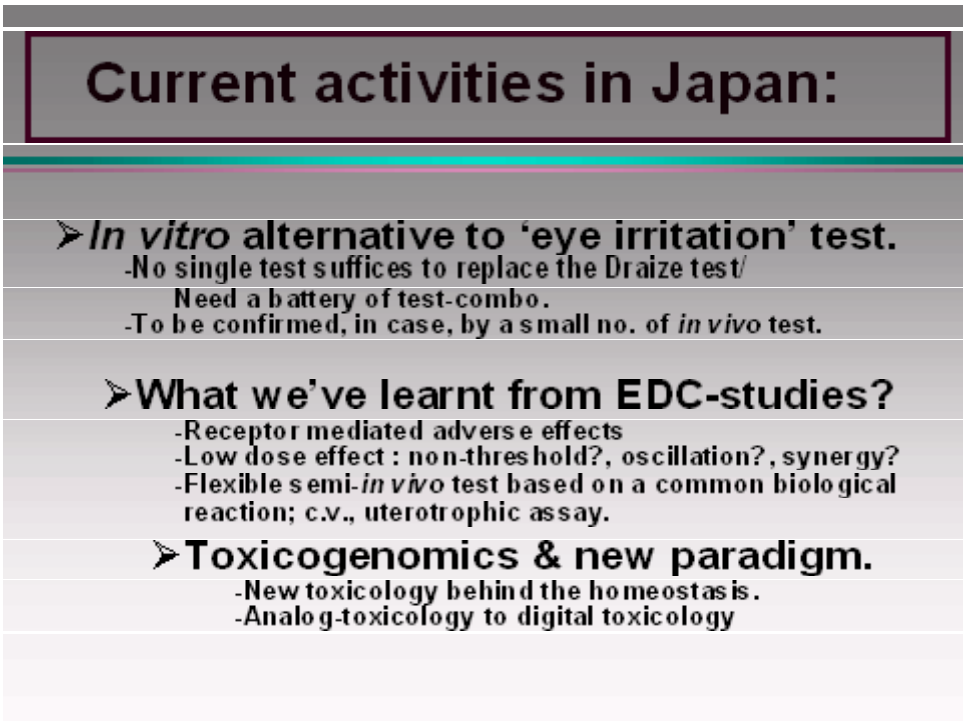
ANNEX 6

Presentation made by: Tohru Inoue, NIHS,MHLW, Japan






What would be the driving-forces to fill the gap?
Eventually, animal study may be replaced by other technology, and disappeared
Gaps to be filled
Animal studies should be replaced by other technology, immediately, hopefully tomorrow!
Basic Driving Forces: Humane animal welfare, economy, politics, etc.
Essential Driving Forces: 1. Science (specifically genome sciences)
2. "l'oiseau bleu"


Current activities in Japan:
In vitro alternative to 'eye irritation' test.
-No single test suffices to replace the Draize test/ Need a battery of test-combo.
-To be confirmed, in case, by a small no. of in vivo test.
What we've learnt from EDC-studies?
-Receptor mediated adverse effects
-Low dose effect : non-threshold?, oscillation?, synergy?
-Flexible semi-in vivo test based on a common biological reaction; c.v., uterotrophic assay.
Toxicogenomics & new paradigm.
-New toxicology behind the homeostasis.
-Analog-toxicology to digital toxicology

Current activities in Japan:

In vitro alternative to 'eye irritation' test

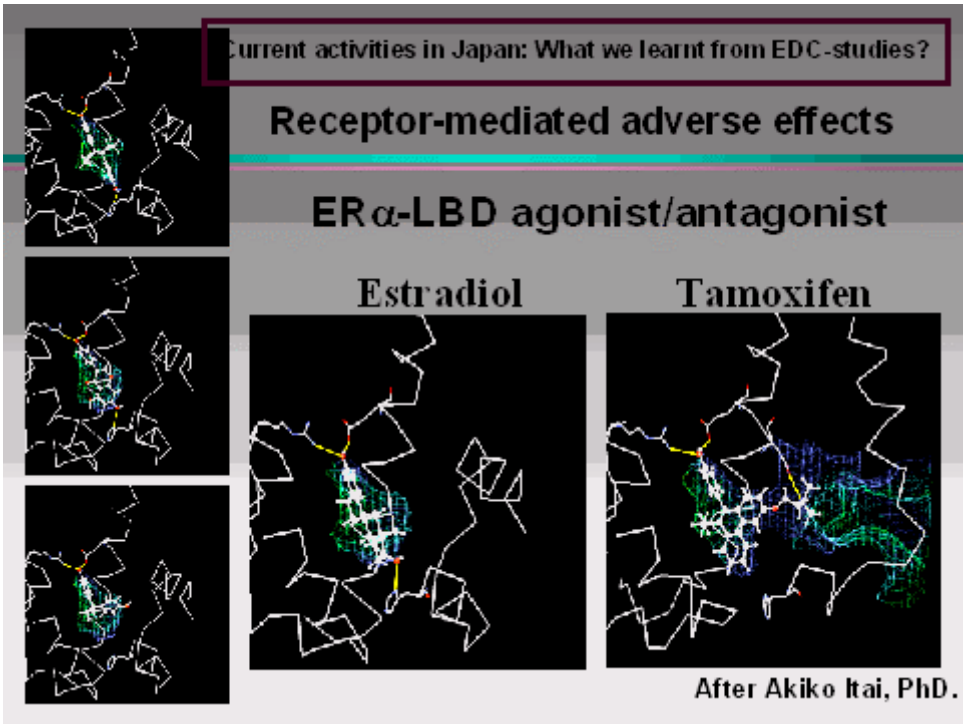
- Alternative in vitro test to the Draize eye irritation test:
- Twelve alternative methods:

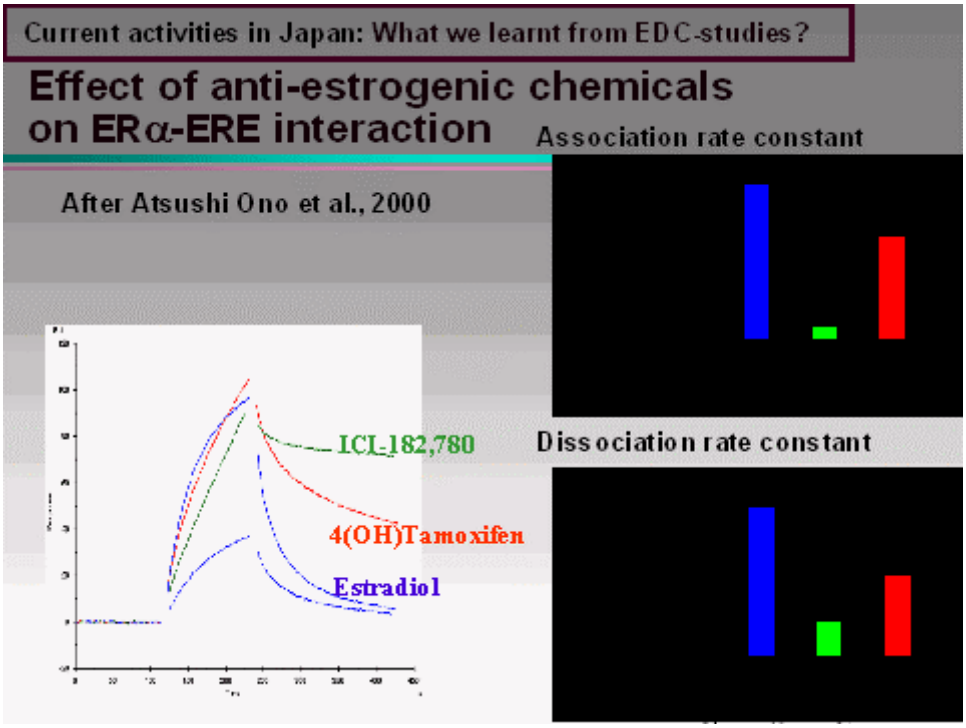
  Chorioallantoic membrane (HET-CAM/CAM-TB), Hemoglobin (HD), Artificial skin models (Skin™/ MATREX™), Normal cells from rabbit cornea (CorneaPack), Cell lines from rabbit cornea (SIRC-CVS/SIRC-NRU), Cell lines from the other mammals (HeLa-MTT/ CHL-CVS), EYTEX™.
- Thirty eight cosmetic ingredients and isotonic sodium chloride solutions:

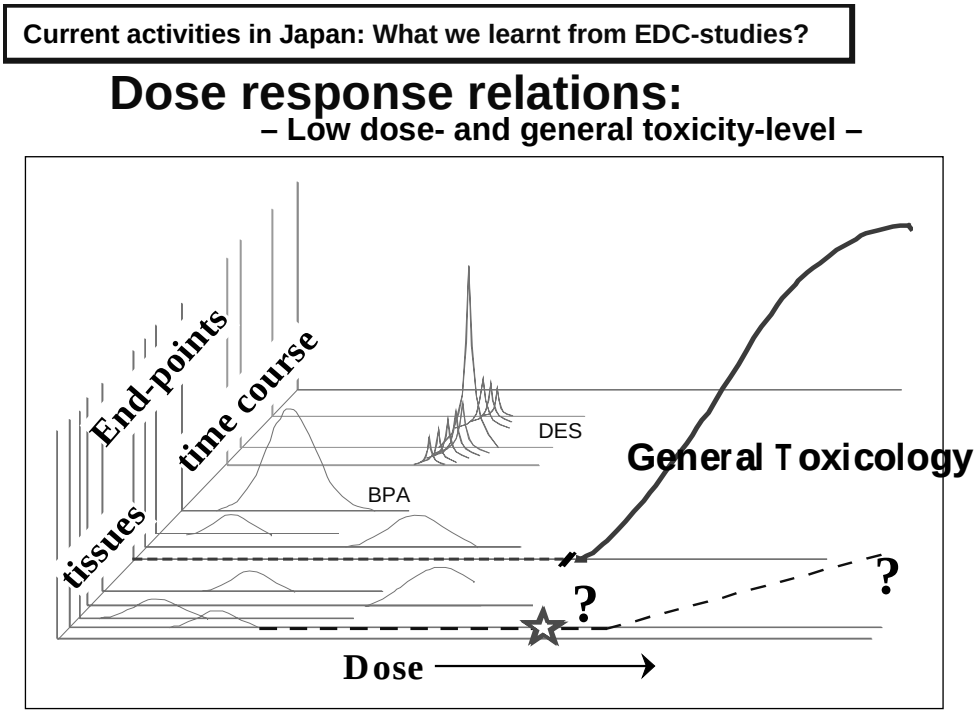
  Isotonic sodium chloride sol/ Polyoxyethylene hydrogenated castor oil / Tween 80 /Polyethyleneglycol monolaurate / Sodium N-lauroyl sarcosinate/ Sodium N-hydro-genated tallow l-glutamate/ Sodium lauryl sulfate/ Sodium polyoxyethylene lauaryl-ether sulfate / Polyoxyethylene octylphenylether (Triton X-100)/ Benzalkonium chloride/ Sucrose fatty acid ester/ Glycerin/ Acid red 92/ Polyoxy-ethylene sorbitan monoleate/ Calcium thoglycolate/ Distearyldimethylammonium chloride/ 2-ethylhexyl p-dimethylamino benzoate/ Cetylpyridinium chloride/ Methyl p-hydroxybenzoate/ Isopropyl myristate/ Polyethylene glycol 400/ Siilicic acid / Benzyl alcohol/ Sodium salicylate/ m-phenylene diamine/ Ethanol / Monoethanolamine/ Triethanolamine / Stearyltrimethylammonium chloride/ Diisopropanolamine/ Potassim laurate/ Cetyltrimethylammonium bromide/ Acetic acid/ Butanol/ Chlorhexidine gluconate solution/ Domiphen bromide/ Lactic acid/ Glycolic acid/ Di(2-ethylhexyl) sodium sulfosuccinate.
- Testing results:

  -No single test suffices Draize / Need a battery of test-combo.

  -To confirm 'a negative', in case, a small no. of *in vivo* test is required.

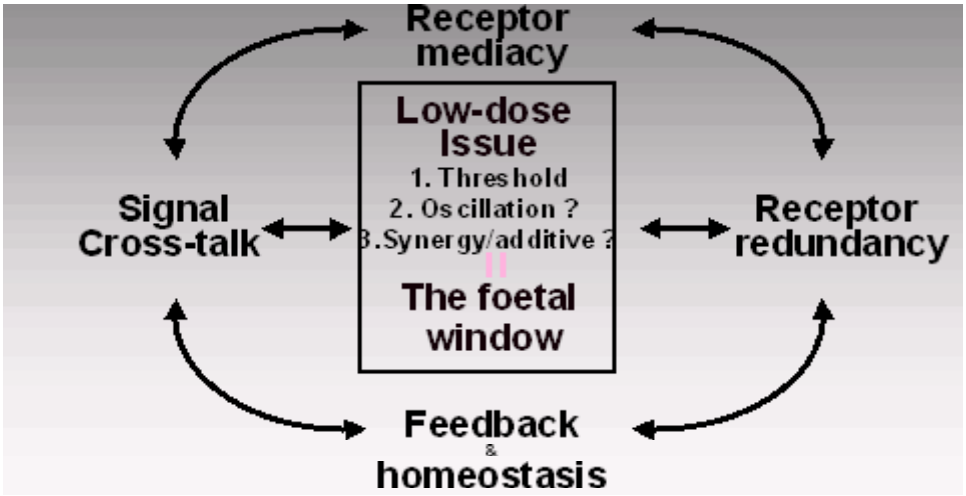



Current activities in Japan: What we learnt from EDC-studies?
Effect of anti-estrogenic chemicals
on ERα-ERE interaction
Association rate constant
After Atsushi Ono et al., 2000
ICI-182,780
4(OH)Tamoxifen
Estradiol
Dissociation rate constant


Current activities in Japan: What we learnt from EDC-studies?
Dose response relations:
– Low dose- and general toxicity-level –
End-points
time course
tissues
DES
BPA
General Toxicology
?
?
Dose

Current activities in Japan: What we learnt from EDC-studies?

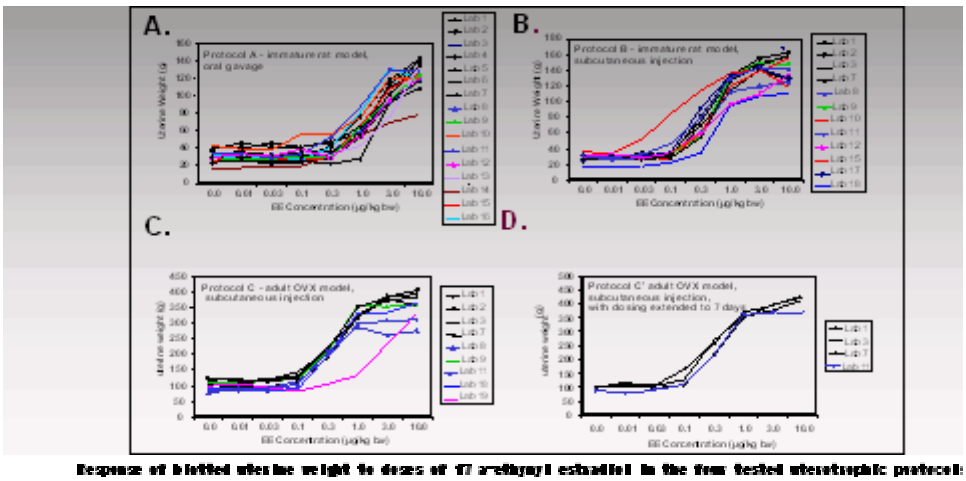
Possible mechanism of the low dose issue



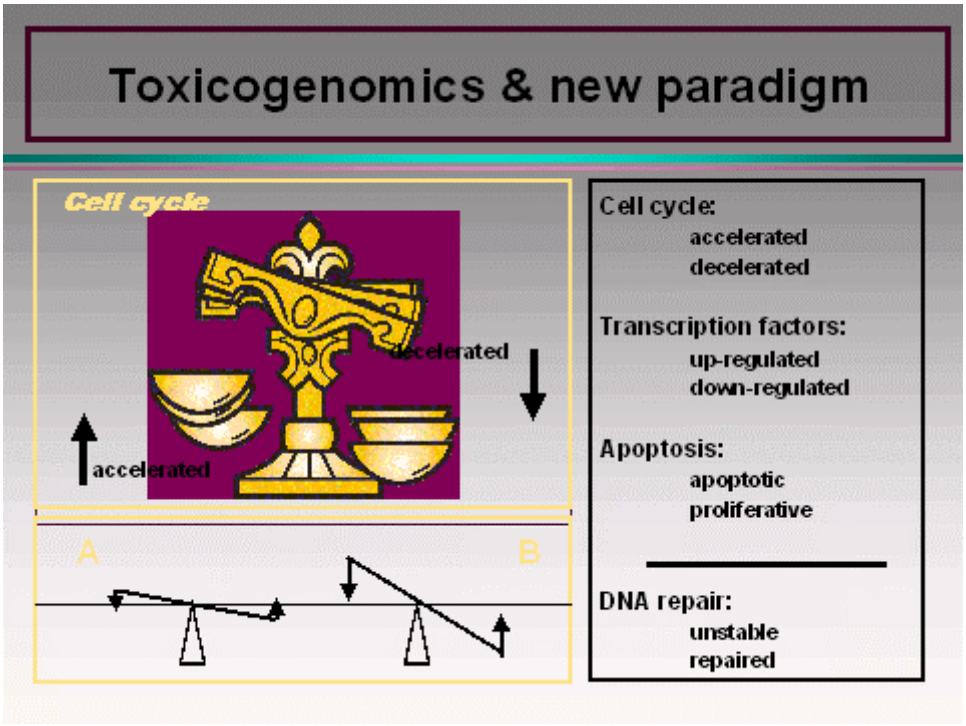
Current activities in Japan: What we learnt from EDC-studies?

Flexible semi-*in vivo* test based on a common biological reaction; uterotrophic assay

A. Immature rat / oral gavages, B. Immature rat / s.c.
C. Adult ovx / s.c., D. Adult ovx / s.c., 7 day-dosing.



Response of blotted uterine weight to doses of 17 α-ethynyl estradiol in the four tested uterotrophic protocols


Toxicogenomics & new paradigm
Cell cycle
decelerated
accelerated
A
B
Cell cycle:
accelerated
decelerated
Transcription factors:
up-regulated
down-regulated
Apoptosis:
apoptotic
proliferative
DNA repair:
unstable
repaired

ANNEX 7

Remarks Prepared for Delivery by

Susan Hazen, Principal Deputy Assistant Administrator
EPA's Office of Prevention, Pesticides and Toxic Substances

OECD Conference on Validation and Regulatory Acceptance
of New and Updated Test Methods in Harzard Assessment

Stockholm
March 6, 2002

DRAFT 2/12/02

Good morning. It's a pleasure to be here. I'd like to thank all of our gracious hosts in Stockholm, especially Environment Minister Larsson and our conference chair Bo Wahlstrom. I am sorry that neither one could be here today, but I wish them both a speedy recovery.

I'd also like to thank other organizers of this important conference, the OECD Secretariat, and especially Herman Koeter, (Principle Administrator for the Test Guidelines Programme) for giving me the opportunity to talk to you today.

The United States is pleased to be working together under the auspices of OECD on yet another important harmonization project. We're all taking another step forward together in our quest to ensure the global safety of chemicals.

This is a time when so many of our efforts in this area are beginning to come to fruition. Two important treaties that address the global threats have been signed and are now being ratified. The first addresses global threats posed by Persistent Organic Compounds; the second establishes procedures for countries to give Prior Informed Consent before allowing importation of 27 toxic industrial chemicals and pesticides.

OECD's Screening Information Data Set (SIDS) Program and industries' International Council of Chemical Associations (ICCA) chemical testing program complement the United States' efforts to provide screening-level data for High Production Volume, or "HPV" Chemicals. These international collaborations ensure that the United States and other developed countries "share the burden" of testing and assessing international HPV chemicals. More importantly, these collaborations ensure that chemicals will not be present in our homes and workplaces without the benefit of basic toxicity testing.

And our harmonization work is an integral part of those efforts – efforts that in years to come will help make the world safer from the risks of toxic chemicals.

We are gathered here to develop considerations and practical guidance for validating test methods that we will all use to ascertain whether and to what degree exposure to test chemicals may pose risks to human health and the environment. This endeavor goes to the very heart of many of the efforts underway. They directly affect us at EPA, and they affect the United States through our interagency efforts on test methods. This effort is central to the work that OECD is performing internationally.

We all have high expectations for our collaboration here, and I sincerely hope our expectations will not only be met but will be exceeded. We've worked together on many harmonization issues and know that we stand on common ground in many areas. Yet, this is a collaboration where each of us brings a unique background and experience. We all have our own perspectives on how to proceed, and it will be our job here to try to ensure broad consensus. In that spirit, I'd like to talk about some of things that the United States is most interested in discussing and resolving.

Our purpose is to build on previous, excellent work in test validation and acceptance, which has become known as the International Solna Principles. Working with other OECD member countries, the United States helped to develop the Solna principles and we fully support them. They are essentially the same as those originally put forward on this side of the Atlantic through the European Centre for the Validation of Alternative Methods (ECVAM) and on my side of the pond by the fifteen federal agencies that make up the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM).

We need to build on the experience gained since the Solna meeting in 1996. Many parties have played roles with test methods, like ZEBET in Germany, the Swiss Center, FRAME, in Britain, the Center for Alternatives to Animal Testing in the U.S. ICCVAM has become a permanent committee in the United States, and it has helped evaluate the Local Lymph Node Assay for contact hypersensitivity and a revised up-and-down procedure for acute oral toxicity. These methods have been accepted by OECD.

The Solna principles provide a strong foundation for the work that lies before us. Our job now is to build in more guidance. Wherever possible we want to end up with a fairly comprehensive set of standard operating procedures that will be helpful to test developers and regulatory authorities. However, we do not want to prescribe any one validation scheme to the exclusion of others. On the one hand, we want to

preserve flexibility; on the other hand, we want to embrace common principles that ensure test methods are scientifically sound and are able to do what is expected when used. We want to know both the strengths and weaknesses of new methods. We want to end up with methods that are both accurate and workable in the real word of limited resources. We want them to be both trustworthy to scientists and trusted by the public.

We also want to embrace new technologies that may be faster and cheaper, but we must be sure that moving away from current methods will not compromise our ability to protect human health and the environment.

In order to achieve those goals, the United States hopes that we will consider some of the principles that we think are of the utmost importance.

There are four major ones:

· Balancing flexibility in the validation process with a need for some degree of specificity;

· Preserving good science;

· Using independent peer review; and,

· Consideration of the 3Rs -- reduction, refinement, and replacement in test method design.

Let's look at each of these in more detail:

1. Flexibility: There should be some level of flexibility in designing validation programs. It should not be strictly uniform across all tests, as one size does not fit all. We need to make important differentiations among the types of tests that we consider. For example, it might be desirable and practical to run hundreds of chemicals through an *in vitro* screen in order to validate it. To do the same for animal protocols is neither practical nor desirable.

It's our hope that the guidance you develop strikes a balance between ensuring the rigor and accuracy of a test with the constraints we face in the real world. In establishing sound guidance, we cannot let the perfect

totally drive out the good. Both positions can prevail. However, we do need to identify what is good, and I urge people here to deliberate hard on what is good enough.

2. Scientific principles: We want a preservation of good scientific principles. After all, validation is a scientific undertaking. Attention needs to be given to identification and evaluation of a specific protocol that will be employed.

There must be demonstration of intra and interlaboratory variability.The method must be shown to be relevant to the health effect being assessed. Attention needs to focus on the number of animals used per dose and per test, the number of chemicals that need to be tested, and the number of laboratories to be used. There must be agreement on a validation plan among authorities before work commences.

3. Peer review: We think citizens will trust regulatory agencies only if their decision making is open and transparent -- and based on sound science. I don't think any agency or country can achieve these requisites for public trust without this type of decisionmaking process.

What do I mean by "open and transparent?" I mean that independent peer review must be part of the process. Peer review members must be independent experts, free of financial and other relevant conflicts of interest and chosen through an open process. Peer review must be conducted in a forum open to the public. The public should be given opportunity to comment on draft materials in writing and/or to speak in the public session.

4. Attention to the 3Rs: As new test methods are developed, we need to consider the 3Rs of animal welfare: refinement, reduction, and replacement wherever scientifically feasible. International principles embody the 3Rs, and many national policies and regulations -- including the United States' require consideration of the 3Rs before animals are used. This means using non-animal methods and approaches instead of animals when this is possible, and using only the minimum number of animals required to obtain valid results. It's also imperative to avoid or mininize animal pain and distress consistent with sound scientific principles. For example at EPA, in the High Production Volume (HPV) program, we have reduced the number of animals by nearly 80percent. In OECD, we have developed tier- testing strategies for both eye and skin irritation, which use structure-activity information, physicochemical properties, in vitro test results and observations in humans and animals before commencing definitive animal testing. Other opportunities abound – with the potential of combining endpoints and studies from multiple methods.

In closing, I would like to thank you all for your efforts. All of us has a stake in the outcome of this meeting. Agreement on additional guidance for validating toxicity tests will help us all to move forward with a general understanding and agreement on this important step. Good luck in this endeavor. Thank you.

ANNEX 8

Presentation made by: Michael Balls, ECVAM, EC

Challenges and Issues Relevant to ECVAM in Relation to the Conference Purpose and Objectives

Michael Balls and Andrew Worth

ECVAM, Institute for Health & Consumer Protection
European Commission Joint Research Centre
Italy

OECD Conference, Stockholm, 6-8 March 2002

THE DUTIES OF ECVAM: COMMISSION COMMUNICATION TO THE COUNCIL AND THE PARLIAMENT, 29 OCTOBER 1991

1. To coordinate the validation of alternative test methods at the European Union level.
2. To act as a focal point for the exchange of information on the development of alternative test methods.
3. To set up, maintain and manage a data base on alternative procedures.
4. To promote dialogue between legislators, industries, biomedical scientists, consumer organisations and animal welfare groups, with a view to the development, validation and international recognition of alternative test methods.
5. To help expand the JRC's role in prenormative research.

WHAT IS AN ALTERNATIVE METHOD ?

Three types of alternative result from the Three Rs concept:

1) **Reduction alternatives** provide a comparable level of information from the use of fewer animals, or more information from the same number of animals

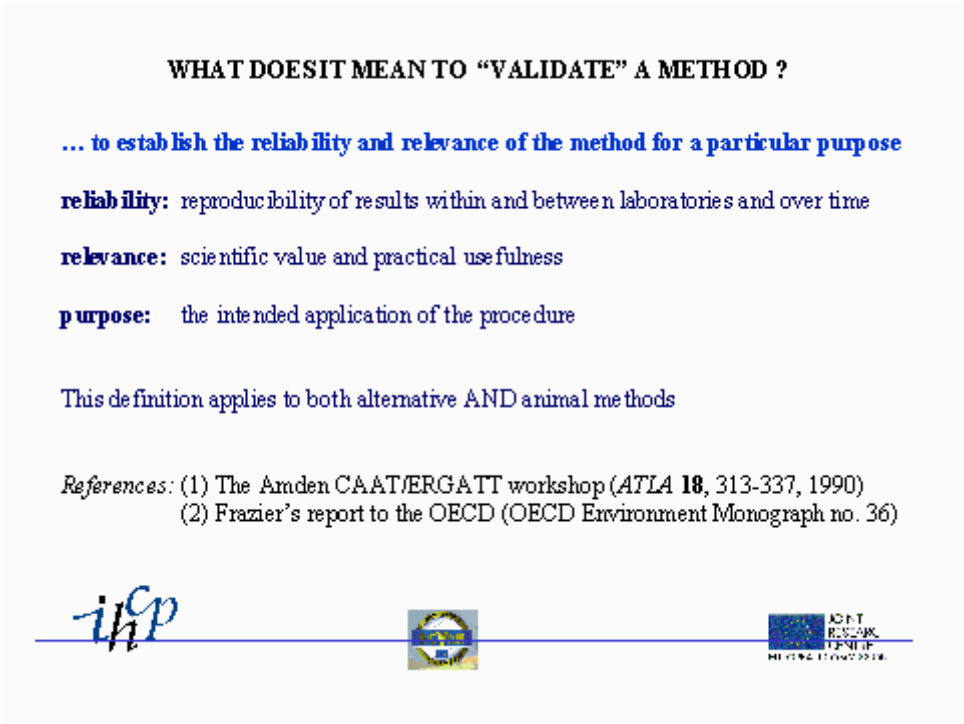
2) **Refinement alternatives** alleviate or minimise potential and unavoidable pain, suffering and distress

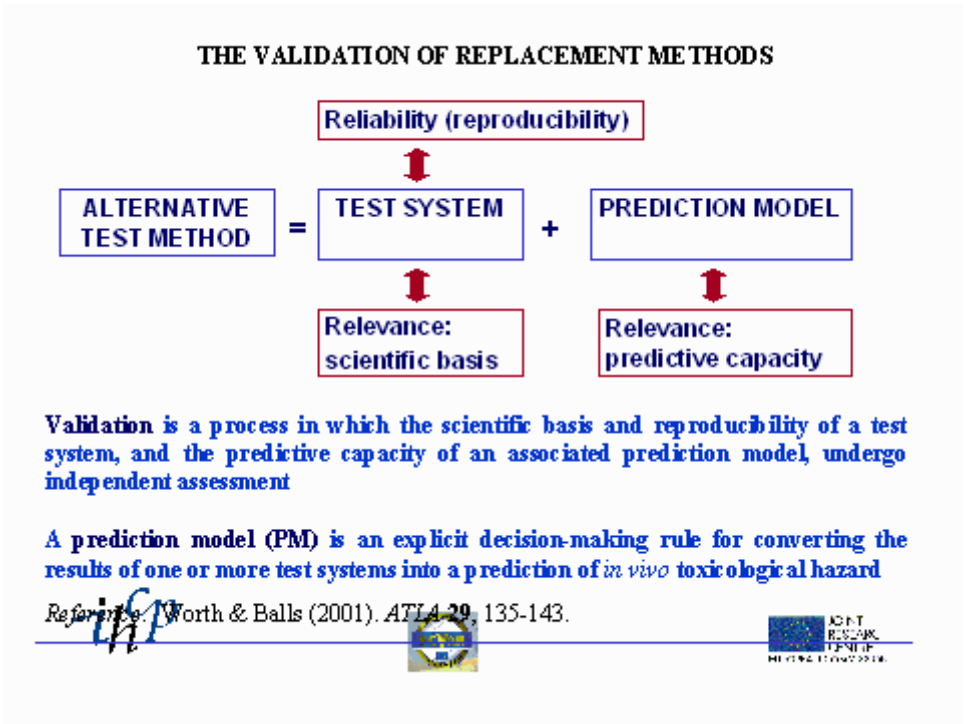
3) **Replacement alternatives** permit a given purpose to be achieved without using living vertebrate animals

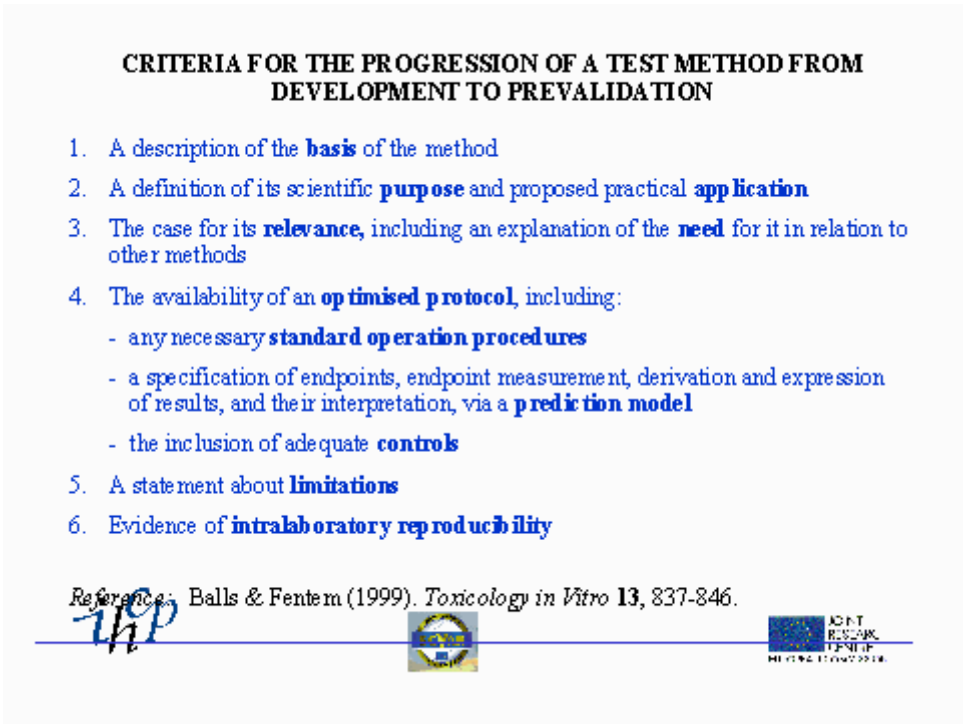
Russell, W.M.S. & Burch, R.L. (1959). *The Principles of Humane Experimental Technique*. Methuen, London.

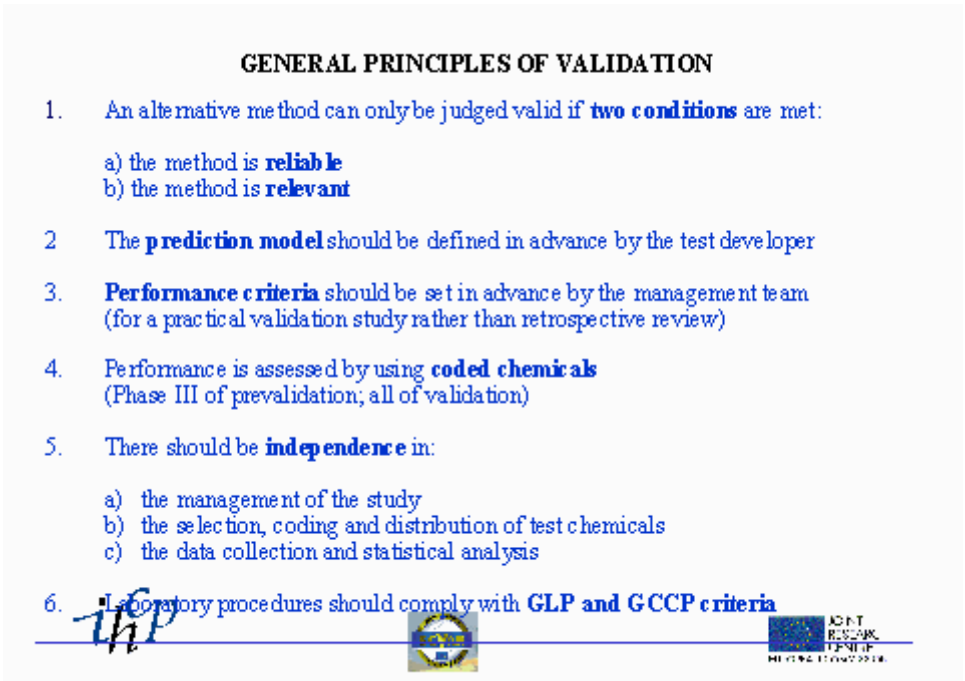
THE ROLE OF ECVAM IN THE EVOLUTION OF REGULATORY TESTS

▪ **Stages:**	1	Research and development	Supporting role of ECVAM
	2	**Prevalidation**	**Leading role of ECVAM**
	3	**Validation**	
	4	Independent assessment	Scientific peer review, ESAC
	5	Regulatory acceptance	Responsibility of other bodies (EU Competent Authorities)
▪ **Timescale: Prevalidation → acceptance**			typically **6 years**


WHAT DOESIT MEAN TO "VALIDATE" A METHOD ?
... to establish the reliability and relevance of the method for a particular purpose
reliability: reproducibility of results within and between laboratories and over time
relevance: scientific value and practical usefulness
purpose: the intended application of the procedure
This definition applies to both alternative AND animal methods
References: (1) The Amden CAAT/ERGATT workshop (ATLA 18, 313-337, 1990)
(2) Frazier's report to the OECD (OECD Environment Monograph no. 36)


THE VALIDATION OF REPLACEMENT METHODS
Reliability (reproducibility)
ALTERNATIVE TEST METHOD
=
TEST SYSTEM
+
PREDICTION MODEL
Relevance: scientific basis
Relevance: predictive capacity
Validation is a process in which the scientific basis and reproducibility of a test system, and the predictive capacity of an associated prediction model, undergo independent assessment
A prediction model (PM) is an explicit decision-making rule for converting the results of one or more test systems into a prediction of in vivo toxicological hazard
Reference: Worth & Balls (2001). ATLA 29, 135-143.


CRITERIA FOR THE PROGRESSION OF A TEST METHOD FROM DEVELOPMENT TO PREVALIDATION
1. A description of the basis of the method
2. A definition of its scientific purpose and proposed practical application
3. The case for its relevance, including an explanation of the need for it in relation to other methods
4. The availability of an optimised protocol, including:
- any necessary standard operation procedures
- a specification of endpoints, endpoint measurement, derivation and expression of results, and their interpretation, via a prediction model
- the inclusion of adequate controls
5. A statement about limitations
6. Evidence of intralaboratory reproducibility
Reference: Balls & Fentem (1999). Toxicology in Vitro 13, 837-846.


GENERAL PRINCIPLES OF VALIDATION
1. An alternative method can only be judged valid if two conditions are met:
a) the method is reliable
b) the method is relevant
2 The prediction model should be defined in advance by the test developer
3. Performance criteria should be set in advance by the management team (for a practical validation study rather than retrospective review)
4. Performance is assessed by using coded chemicals (Phase III of prevalidation; all of validation)
5. There should be independence in:
a) the management of the study
b) the selection, coding and distribution of test chemicals
c) the data collection and statistical analysis
6. Laboratory procedures should comply with GLP and GCCP criteria

CRITERIA FOR PREDICTION MODELS

For a PM to be considered adequate, it should:

1) be associated with one or more specific protocols

2) be associated with a clearly defined pharmacotoxicological endpoint

3) have its limitations clearly defined

4) be associated with an indication of the accuracy of its predictions

For a PM to be considered valid, it should:

1) be assessed with independent data (i.e. the test set should be different to the training set)

2) meet or exceed the criteria for predictive capacity defined by the Management Team of a validation study

Bruner *et al.* (1996). *Toxicology in Vitro* **10**, 479-501.
Worth & Balls (2001). *ATLA* **29**, 135-143.

EXAMPLES OF PREDICTION MODELS

1) PM for skin corrosion potential based on pH measurements

If the pH of a 10% solution of a substance ≤ 2, or if pH ≥ 11.5, classify as a corrosive.

2) PM for skin corrosion potential based on *in vitro* (EPISKIN) data

If the viability of EPISKIN < 35% after treatment for 3 min, classify as a severe corrosive (EU risk phrase R35);

Otherwise if viability < 35% after treatment for 60 min or 240 min, classify as corrosive (EU risk phrase R34);

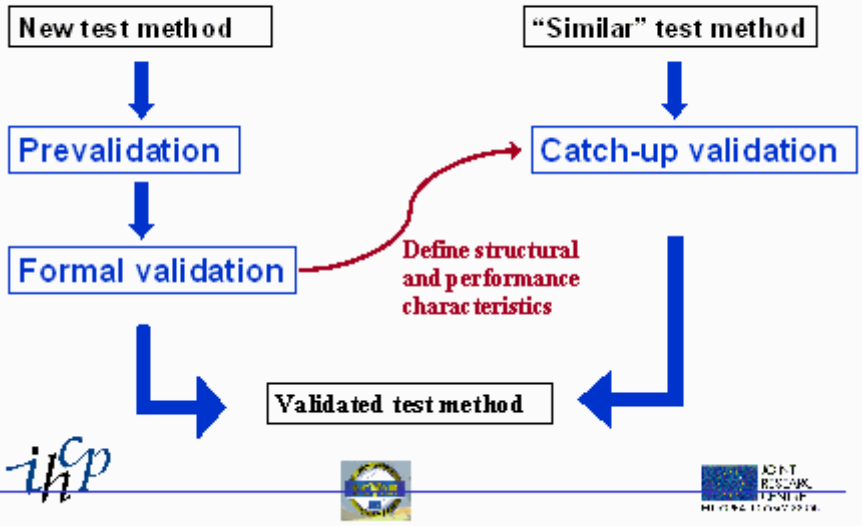
Otherwise, if the viability of EPISKIN ≥ 35% after treatment for 240 min, classify as non-corrosive.

References:

EC (2000). Commission Directive *2000/33/EC. Official Journal of the European Communities* **L136A**, 90-97.

Fentem *et al.* (1998) *Toxicology in Vitro* **12**, 483-524.

OVERVIEW OF THE ECVAM VALIDATION PROCESS



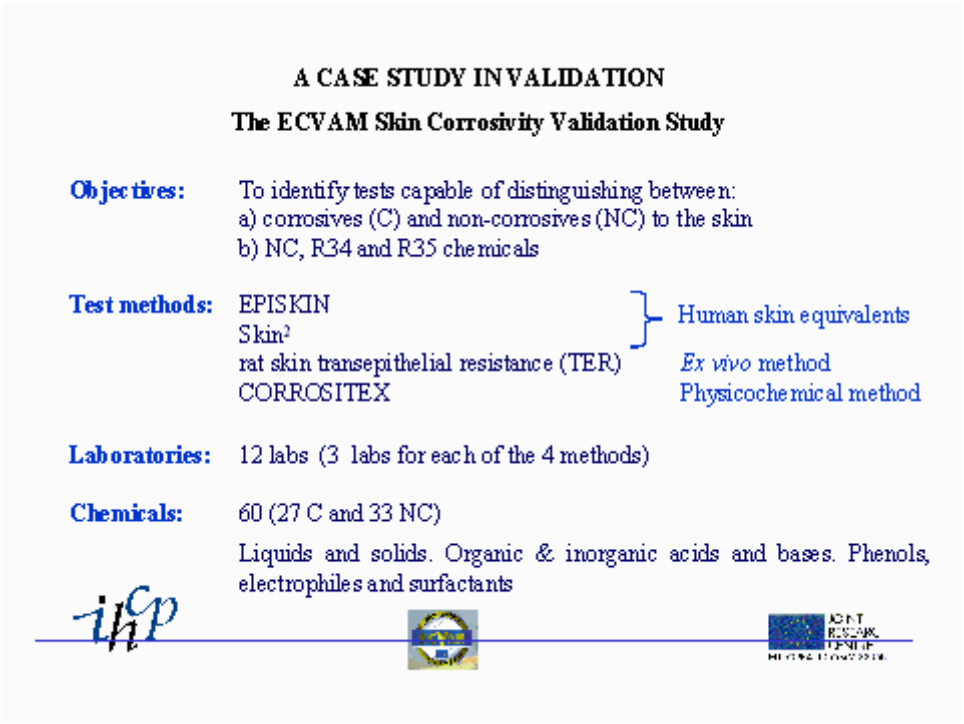
TYPES OF VALIDATION STUDIES

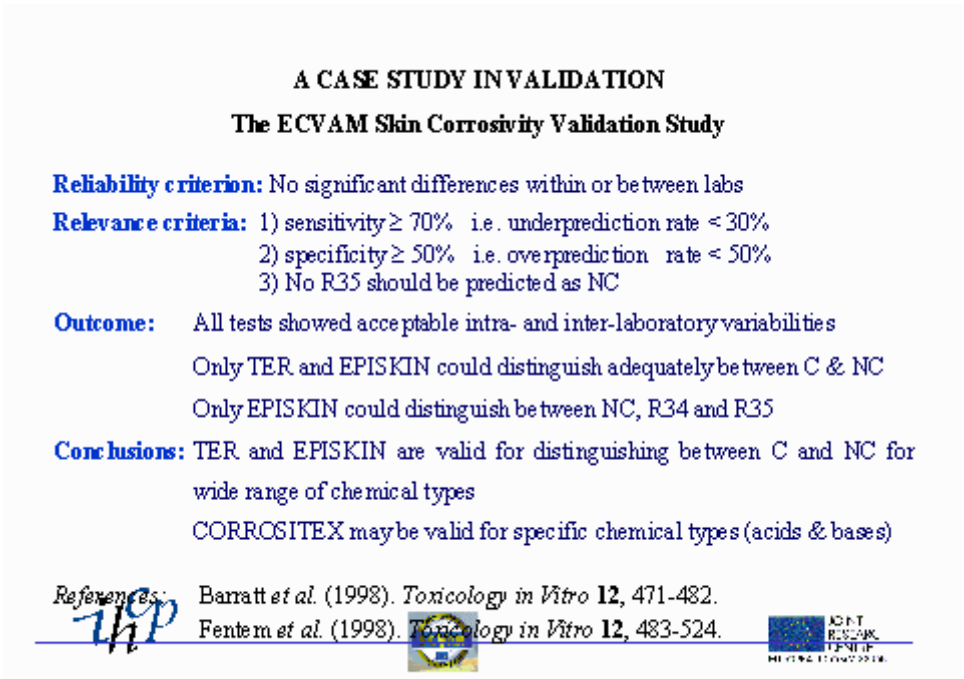
A PREVALIDATION STUDY is a small-scale inter-laboratory study, carried out to ensure that the protocol of a test method is sufficiently optimised and standardised for inclusion in a formal validation study

Phase I	**Refinement**	Laboratory 1 "lead laboratory"
Phase II	**Transfer**	Laboratory 1 → Laboratory 2
Phase III	**Performance**	Laboratories 1, 2 & 3

A VALIDATION STUDY is a large-scale inter-laboratory study, designed to assess the reliability and relevance of an optimised method for a particular purpose

A CATCH-UP VALIDATION STUDY is a prevalidation study in which the structural and performance criteria of a method are compared with those of a similar method, which has already undergone formal validation and has been accepted as scientifically valid


A CASE STUDY INVALIDATION
The ECVAM Skin Corrosivity Validation Study
Objectives:
To identify tests capable of distinguishing between:
a) corrosives (C) and non-corrosives (NC) to the skin
b) NC, R34 and R35 chemicals
Test methods:
EPISKIN
Skin²
rat skin transepithelial resistance (TER)
CORROSITEX
Human skin equivalents
Ex vivo method
Physicochemical method
Laboratories:
12 labs (3 labs for each of the 4 methods)
Chemicals:
60 (27 C and 33 NC)
Liquids and solids. Organic & inorganic acids and bases. Phenols, electrophiles and surfactants


A CASE STUDY INVALIDATION
The ECVAM Skin Corrosivity Validation Study
Reliability criterion: No significant differences within or between labs
Relevance criteria: 1) sensitivity ≥ 70% i.e. underprediction rate < 30%
2) specificity ≥ 50% i.e. overprediction rate < 50%
3) No R35 should be predicted as NC
Outcome:
All tests showed acceptable intra- and inter-laboratory variabilities
Only TER and EPISKIN could distinguish adequately between C & NC
Only EPISKIN could distinguish between NC, R34 and R35
Conclusions:
TER and EPISKIN are valid for distinguishing between C and NC for wide range of chemical types
CORROSITEX may be valid for specific chemical types (acids & bases)
References:
Barratt et al. (1998). Toxicology in Vitro 12, 471-482.
Fentem et al. (1998). Toxicology in Vitro 12, 483-524.

A CASE STUDY IN PREVALIDATION / CATCH-UP VALIDATION

The EpiDerm Method for Skin Corrosion Potential

- EPISKIN and EpiDerm are both human skin equivalents
- EPISKIN was scientifically validated in the Skin Corrosivity Validation Study
- The scientific validity of EpiDerm was assessed in an ECVAM catch-up study Liebsch *et al.* (2000) *ATLA* **28**, 371-401.

- **Study Design:** 3 labs, 24 chemicals from SCVS (12 C and 12 NC)
- **Results:** No significant inter-laboratory differences ⇒ **reliable**
  Sensitivity = 88%; specificity = 86 % ⇒ **relevant**
- **Conclusion:** EpiDerm is valid for predicting the corrosion potential (C/NC) of a wide range of chemicals (including neutral organics, acids, bases, electrophiles and phenols)

THE ROLE OF EUROPEAN COMMISSION SERVICES IN THE EVOLUTION OF REGULATORY TESTS

Several services of the Commission are represented on the ESAC:

1) Joint Research Centre (JRC; ECVAM, the ECB and the IHCP)
   - provide scientific and technical advice to the JRC's customer DGs

2) Environment DG (Unit C.3)
   - manages legislation relating to the protection of laboratory animals (*Directive 86/806/EEC*)
   - manages legislation relating to the testing of chemicals (*Directive 67/548/EEC*)

3) Enterprise DG (Unit F.3)
   - manages legislation relating to products (*Directive 88/379/EEC*) and cosmetics (*Directive 76/768/EEC*)

4) Health & Consumer Protection DG (Unit C.2)
   - coordinates meetings of the SCC-NFP

5) Research DG (Unit E.4)
   - provides funding for the development of alternative methods

THE ROLE OF THE ECVAM SCIENTIFIC ADVISORY COMMITTEE (ESAC) IN THE EVOLUTION OF REGULATORY TESTS

The ECVAM Scientific Advisory Committee (ESAC) comprises representatives of the Member States, of the European chemical, cosmetic and pharmaceutical industry associations, of academic toxicology and animal welfare organisations, and of several Commission services

The ESAC has the following tasks:

1) to advise ECVAM on the development, validation and acceptance of alternative methods

2) to promote the activities of ECVAM in the EU Member States

3) to promote the acceptance of validated methods in the EU Member States

The ESAC makes a formal recommendation on the scientific validity of an alternative method by issuing a statement on the applicability of the method for a particular purpose

Reference: Worth & Balls (2001). *ATLA* **29**, 525-535.

COMPOSITION OF THE ESAC

Representation		Number of members
European Union	Member States	15 (1 from each Member State)
Industry organisations	COLIPA	1
	ECETOC	1
	EFPIA	1
Academic organisations	ERGATT	1
Animal welfare organisations	Eurogroup	2
European Commission services	JRC	5 (3 ECVAM, 1 ECB, 1 IHCP)
	DG ENV	1 (Unit C.3)
	DG ENTR	1 (Unit F.3)
	DG SANCO	1 (Unit C.2)
	DG RES	1 (Unit E.4)

COMPLETED PREVALIDATION AND VALIDATION STUDIES - Chemicals

1. Nine methods for eye irritation (EC/HO Eye Irritation Study)
   (*Toxicology in Vitro* **9**, 971-929, 1995)
2. Five methods for skin corrosivity:
   a) rat skin TER, CORROSITEX, EPISKIN and Skin^2 (*ATLA* **23**, 291-255, 1995)
   b) EpiDerm - Catch-up validation (*ATLA* **28**, 371-401, 2000)
3. Five methods for skin irritation:
   a) EpiDerm, EPISKIN, PREDISKIN, the non-perfused pig ear test
   (*Toxicology in Vitro* **15**, 57-93, 2001)
   b) Skin Integrity Function Test (SIFT)
4. For phototoxic potential: the 3T3 Neutral Red Uptake (NRU) test
   EU/COLIPA study (*Toxicology in Vitro* **12**, 305-327, 1998)
   Applicability to UV filters (*ATLA* **26**, 679-708, 1998)
5. Three methods for embryotoxic potential:
   Embryonic stem cell test, micromass test and rat embryo culture test
   (*Toxicology in Vitro* **15**, 57-93, 2001)
6. For acute neutropenia:
   Colony Forming Unit-Granulocyte/Macrophage (CFU-GM) assay

METHODS ENDORSED BY THE ESAC - Chemicals

ESAC Statement	Date
3T3 NRU phototoxicity test	03-11-97
EPISKIN skin corrosivity test	03-04-98
Rat TER skin corrosivity test	03-04-98
Application of the 3T3 NRU phototoxicity test to UV filter chemicals	20-05-98
Local lymph node assay for skin sensitisation	21-03-00
EpiDerm skin corrosivity test	23-03-00
CORROSITEX skin corrosivity test	06-12-00
Embryonic stem cell test for embryotoxicity	17-10-01
Whole-embryo culture test for embryotoxicity	17-10-01
Micromass test for embryotoxicity	17-10-01

REGULATORY ACCEPTANCE OF ALTERNATIVE METHODS - Chemicals

- **On 4 February 2000, the EU Competent Authorities** for *Directive 67/548/EEC* (on the Classification, Packaging and Labelling of Dangerous Substances) **accepted three** *in vitro* **methods** as replacement methods for the toxicity testing of chemical substances:

1. The 3T3 NRU method for phototoxic potential
2. The rat skin transcutaneous electrical resistance (TER) method for skin corrosion potential
3. Human skin equivalents (that meet certain criteria) for skin corrosion potential (EPISKIN & EpiDerm)

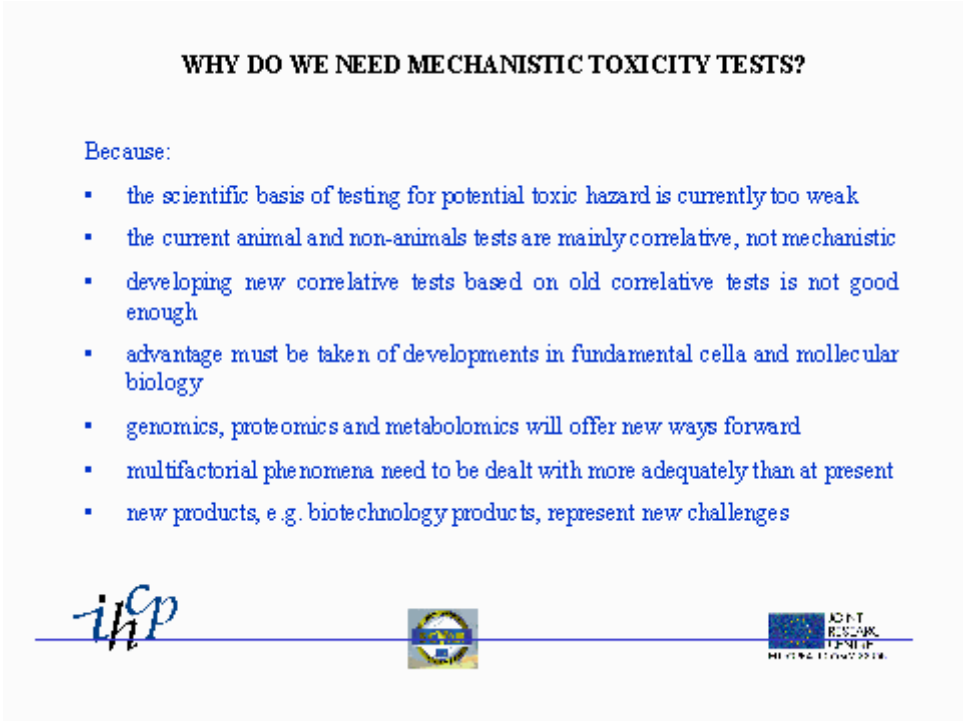
- These methods have been incorporated into **Annex V** of *Directive 67/548/EEC*

- **Test guidelines** for these methods have been submitted by DG ENV to the OECD Secretariat, and are now **under consideration by OECD Member Countries**

SOME UNAVOIDABLE CHOICES IN VALIDATION STUDIES

- number of tests (as low as possible)
- number of endpoints (must be complementary - avoid duplication)
- number of test items (as high as possible)
- number of laboratories (as low as possible, maximum = 4)


NEW APPROACHES TO VALIDATION ARE NECESSARY
Because:
animal test data cannot represent the gold standard to be met by alternative tests
such data are usually lacking, except for a few examples
prospective predictions must replace retrospective predictions
tests are needed for target organ toxicity and chronic toxicity, not just topical toxicity
new kinds of products have to be tested, e.g. biotechnology products
tests based on mechanistic understanding will be used to test for toxicity for which a sufficient mechanistic understanding is available
ihcp

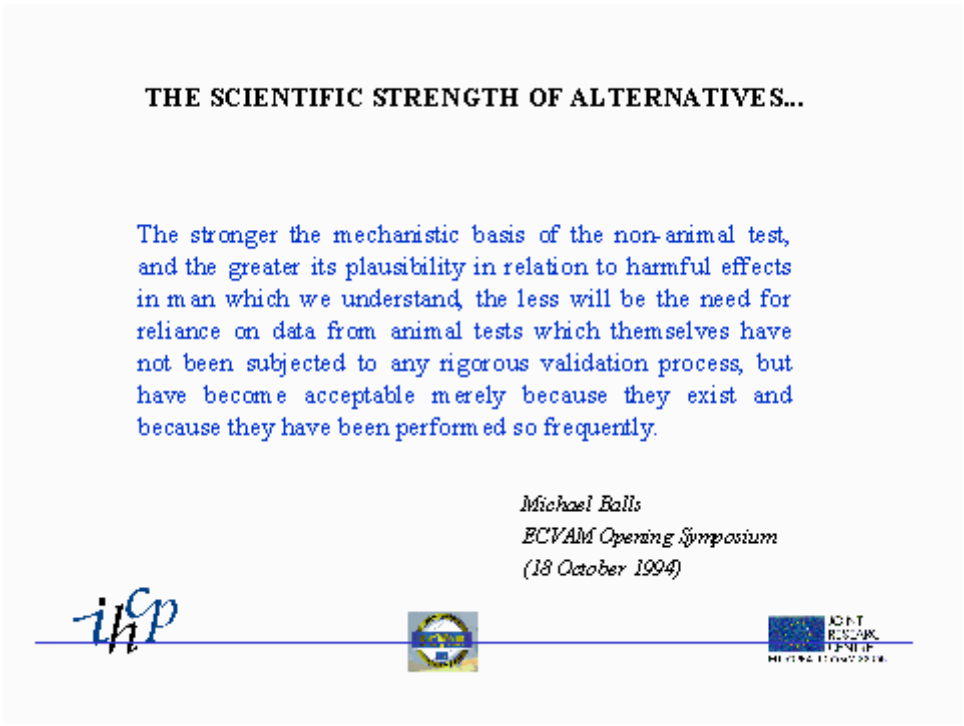

WHY DO WE NEED MECHANISTIC TOXICITY TESTS?
Because:
the scientific basis of testing for potential toxic hazard is currently too weak
the current animal and non-animals tests are mainly correlative, not mechanistic
developing new correlative tests based on old correlative tests is not good enough
advantage must be taken of developments in fundamental cella and mollecular biology
genomics, proteomics and metabolomics will offer new ways forward
multifactorial phenomena need to be dealt with more adequately than at present
new products, e.g. biotechnology products, represent new challenges
ihcp

FACTORS AFFECTING THE PROGRESS OF ALTERNATIVE METHODS

- development of candidate new test methods for pre-validation
- provision of adequate resources for funding development and validation studies
- availability of trained personnel to design and manage pre-validation studies
- availability of experienced laboratories to take part in pre-validation studies
- availability of sufficient reference standard compounds
- quality of independent evaluation of validation studies
- willingness of industries and regulatory authorities to accept scientifically valid alternative test procedures and testing strategies
- strength of insistence on the application of laws which require that animal procedures must be replaced when such alternatives are available.

Areas which Could have Validated Replacement Alternative Methods within Five Years
(if the problems currently limiting progress are solved)

1. Eye irritation
2. Skin irritation
3. Skin sensitisation
4. Skin penetration
5. Nephrotoxicity
6. Hepatoxicity
7. Neurotoxicity
8. The blood-brain barrier
9. Acute systemic toxicity
10. Chemical carcinogenesis


THE SCIENTIFIC STRENGTH OF ALTERNATIVES...
The stronger the mechanistic basis of the non-animal test, and the greater its plausibility in relation to harmful effects in man which we understand, the less will be the need for reliance on data from animal tests which themselves have not been subjected to any rigorous validation process, but have become acceptable merely because they exist and because they have been performed so frequently.
Michael Balls
ECVAM Opening Symposium
(18 October 1994)


A CHALLENGE TO THE REGULATORS...
Meanwhile, can the OECD and all other agencies involved in the current practice of regulatory toxicology reassure us - indeed, prove to us - that their standards for the acceptance of new animal test procedures are not dramatically less stringent than those which will be applied to non-animal tests?
Michael Balls
ECVAM Opening Symposium
(18 October 1994)

OUR LEGAL STRENGTH...

We need not be hesitant in demanding that scientifically valid and feasible non-animal methods and testing strategies should be incorporated into regulatory testing guidelines. *Directive 86/609/EEC* specifically requires that "an experiment [on a laboratory animal] should not be performed, if another scientifically satisfactory method, not entailing the use of an animal, is reasonably and practicably available". The law is on our side.

Michael Balls
ECVAM Opening Symposium
(18 October 1994)

CONCLUDING REMARKS

- **Validation** procedures are still evolving. However, the principles and procedures established so far by ECVAM and other organisations **have proven successful**
- **Good progress** regarding the validation of replacement methods for skin corrosion, and of refinement methods for skin sensitisation
- **Progress still needed** regarding the validation of replacement alternatives for
  a) skin irritation, skin sensitisation, and eye irritation
  b) systemic toxicity
  c) chronic toxicity
- **More emphasis needed on** the validation of methods in the context of **testing strategies**, rather than as stand-alone alternatives
- **ECVAM needs test developers to submit their methods for prevalidation and validation**

ANNEX 9

William Stokes, ICCVAM, USA

Validation and Acceptance of New Test Methods: ICCVAM Experience and Perspectives

William S. Stokes, D.V.M., Diplomate ACLAM

National Toxicology Program

National Institute of Environmental Health Sciences

Research Triangle Park, North Carolina, USA

OECD Conference on Validation and Acceptance of New Test Methods
Stockholm, Sweden
March 6-8, 2002

Outline

- Introduction and Overview of ICCVAM
- Validation and Acceptance Criteria
- ICCVAM Evaluation Process
- ICCVAM Test Method Evaluations
- Validation and Acceptance Issues
- Future Opportunities and Challenges

History of the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)

1993 Public Law 103-43 directed NIEHS to:

- Develop and validate test methods for acute and chronic safety testing, including alternative methods that can reduce or eliminate the use of animals
- Establish criteria for validation and regulatory acceptance
- Develop a process for regulatory acceptance of scientifically validated methods

History of ICCVAM

1994 *ad hoc* ICCVAM: 15 Federal regulatory and research agencies

1995 Draft report of the *ad hoc* ICCVAM

1997 Final report of the *ad hoc* ICCVAM

1997 ICCVAM established

- Replaced *ad hoc* ICCVAM
- Implemented NIEHS directives: P.L. 103-43

2000 ICCVAM Authorization Act of 2000: P.L. 106-545

- Established ICCVAM as a permanent committee

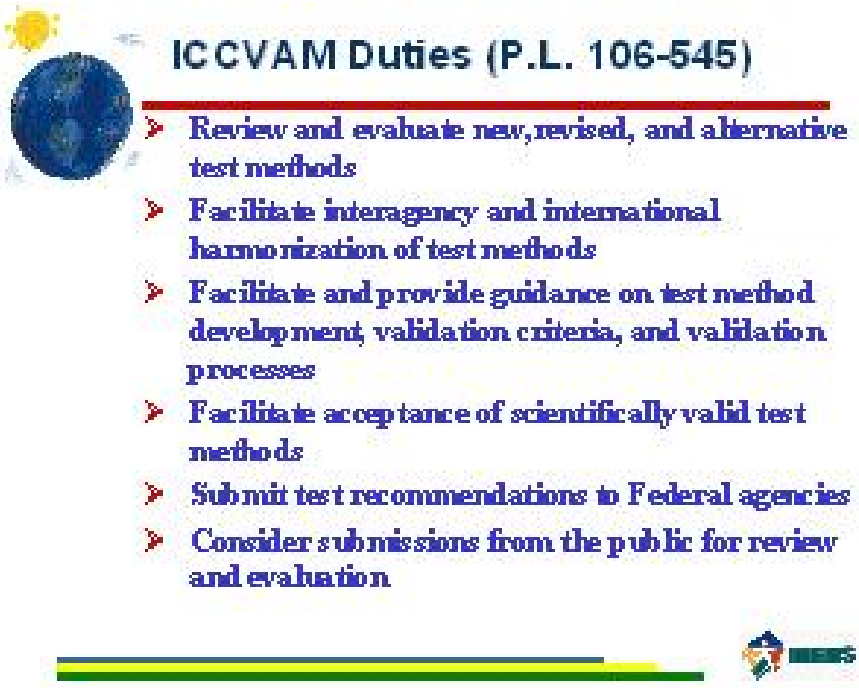
Interagency Coordinating Committee on the Validation of Alternative Methods

Regulatory Research	Non-Regulatory Research
*Consumer Product Safety Commission	*Agency for Toxic Substances and Disease Registry
Department of Agriculture	Department of Defense
Department of Interior	Department of Energy
Department of Transportation	*National Cancer Institute
*Environmental Protection Agency	*National Institute of Environmental Health Sciences
*Food and Drug Administration	*National Institute for Occupational Safety and Health
*Occupational Safety and Health Administration	National Library of Medicine
	*National Institutes of Health, OD

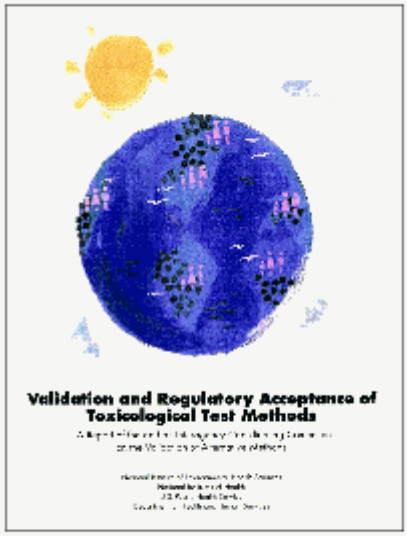
*ITI Executive Committee

Purposes of ICCVAM (P.L. 106-545)

- Increase the efficiency and effectiveness of Federal agency test method review
- Eliminate unnecessary duplicative efforts and share experiences between Federal regulatory agencies
- Optimize utilization of scientific expertise outside the Federal government
- Ensure that new and revised test methods are validated to meet the needs of Federal agencies
- Reduce, refine, or replace the use of animals in testing, where feasible


ICCVAM Duties (P.L. 106-545)
Review and evaluate new, revised, and alternative test methods
Facilitate interagency and international harmonization of test methods
Facilitate and provide guidance on test method development, validation criteria, and validation processes
Facilitate acceptance of scientifically valid test methods
Submit test recommendations to Federal agencies
Consider submissions from the public for review and evaluation


National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)
Located at NIEHS
Functions
Operational/technical support
—ICCVAM
—Test method peer reviews and workshops
—Scientific Advisory Committee
Communication with stakeholders
Partnerships with stakeholders
http://iccvam.niehs.nih.gov



Validation and Regulatory Acceptance of Toxicological Test Methods[1]

- Prepared by the ad hoc ICCVAM
  - Broad stakeholder involvement
  - 1995 - Draft; 1997 - Final
- Provides:
  - Criteria for validation and regulatory acceptance
  - Process for regulatory acceptance
- http://iccvam.niehs.nih.gov/validate.pdf

[1]Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods; NIH Pub. No. 97-3981, 1997, NIEHS, Research Triangle Park, NC. http://iccvam.niehs.nih.gov/docs/guidelines/validatep.pdf

NIEHS

What is Validation?

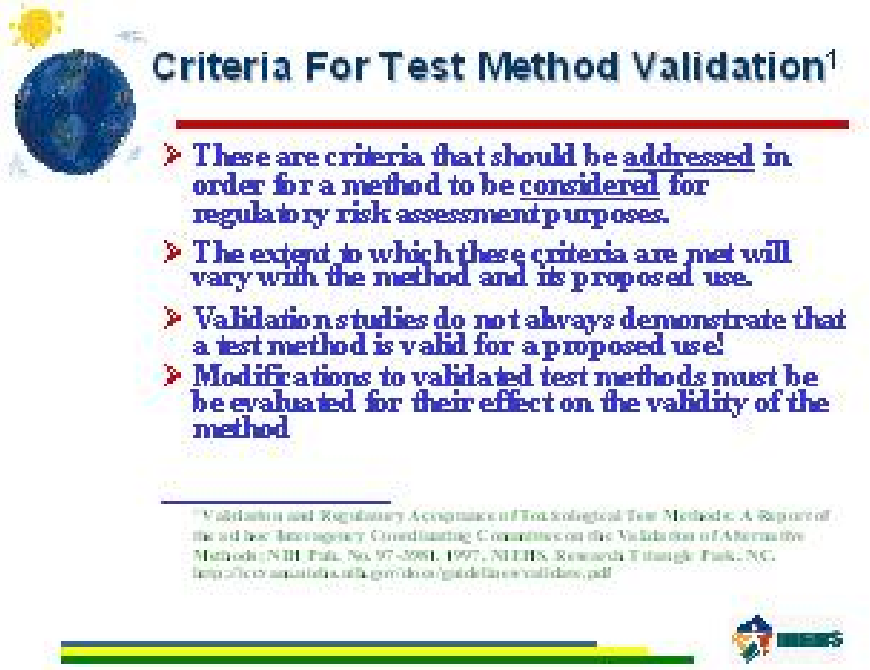
- A determination of the usefulness and limitations of a test method for a specific purpose
- The process by which the *reliability* and *relevance* of a test method are established for a specific purpose.

-Reliability: A measure of the extent to which a test can be performed reproducibly within and among laboratories over time.

-Relevance: The extent to which a test method will correctly predict or measure the biological effect of interest.

NIEHS


Regulatory Acceptance Consideration [1]
Adequate validation is a prerequisite for regulatory acceptance consideration
Will the proposed test method generate data that will provide a comparable or better level of protection of human health or the environment than the current method or approach?


Criteria For Test Method Validation[1]
These are criteria that should be addressed in order for a method to be considered for regulatory risk assessment purposes.
The extent to which these criteria are met will vary with the method and its proposed use.
Validation studies do not always demonstrate that a test method is valid for a proposed use!
Modifications to validated test methods must be be evaluated for their effect on the validity of the method

Criteria For Test Method Validation[1]

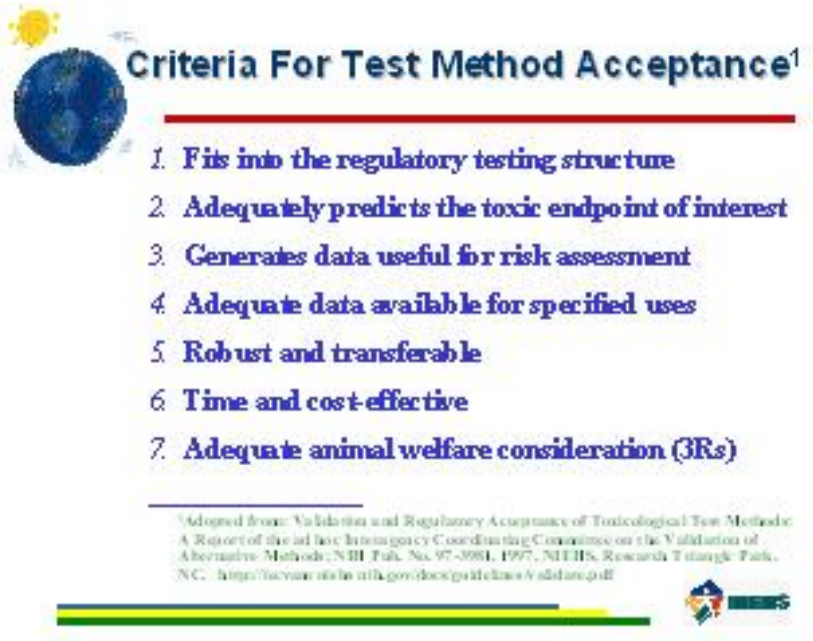
- These are criteria that *should* be addressed in order for a method to be considered for regulatory risk assessment purposes.
- The extent to which these criteria are met will vary with the method and its proposed use.
- Validation *studies* do not always demonstrate that a test method is valid for a proposed use!
- Modifications to validated test methods must be be evaluated for their effect on the validity of the method

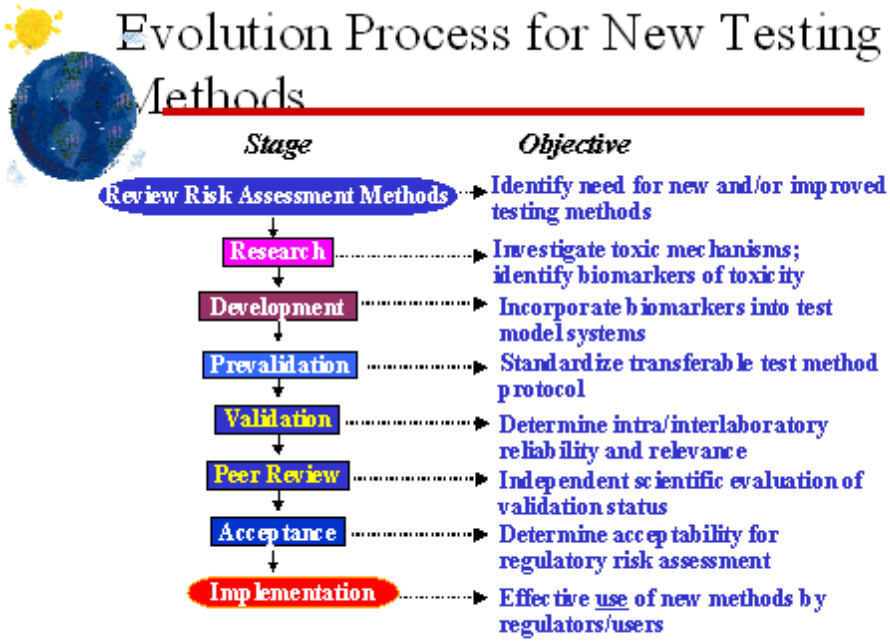
[1] Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods; NIH Pub. No. 97-3981, 1997, NIEHS, Research Triangle Park, NC. http://iccvam.niehs.nih.gov/docs/guidelines/validate.pdf

Criteria for Test Method Validation[1]

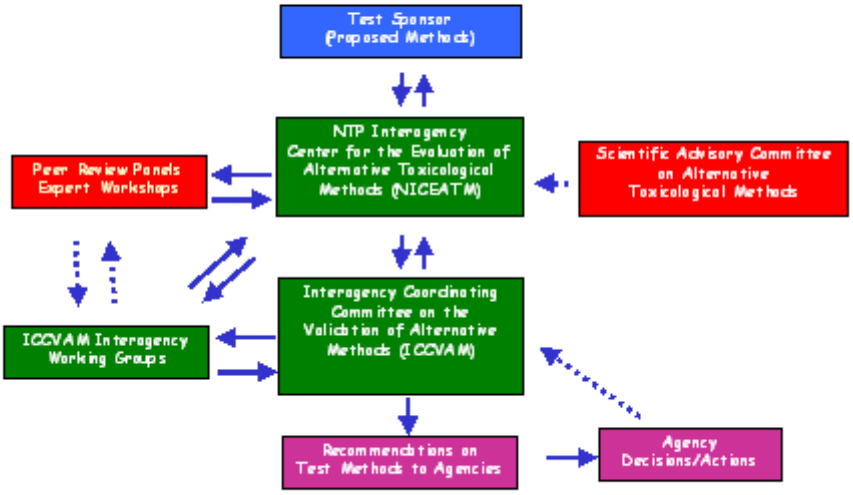
1. Clear *s*tatement of proposed use
2. Biological basis/relationship to effect of interest
3. Formal detailed protocol and SOPs
4. Reliability assessed
5. Relevance assessed
6. Advantages and limitations described
7. All data available for review
8. Data quality: *Ideally GLPs*
9. Independent scientific peer review

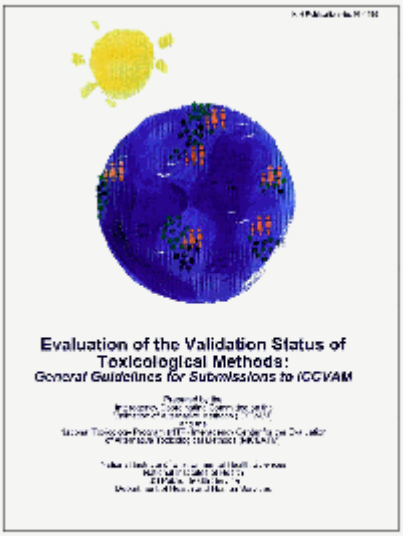
[1] Adapted from: Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods; NIH Pub. No. 97-3981, 1997, NIEHS, Research Triangle Park, NC. http://iccvam.niehs.nih.gov/docs/guidelines/validate.pdf


Criteria For Test Method Acceptance[1]
1. Fits into the regulatory testing structure
2. Adequately predicts the toxic endpoint of interest
3. Generates data useful for risk assessment
4. Adequate data available for specified uses
5. Robust and transferable
6. Time and cost-effective
7. Adequate animal welfare consideration (3Rs)


Evolution Process for New Testing Methods
Stage
Objective
Review Risk Assessment Methods
Identify need for new and/or improved testing methods
Research
Investigate toxic mechanisms; identify biomarkers of toxicity
Development
Incorporate biomarkers into test model systems
Prevalidation
Standardize transferable test method protocol
Validation
Determine intra/interlaboratory reliability and relevance
Peer Review
Independent scientific evaluation of validation status
Acceptance
Determine acceptability for regulatory risk assessment
Implementation
Effective use of new methods by regulators/users

ICCVAM Test Method Evaluation Process

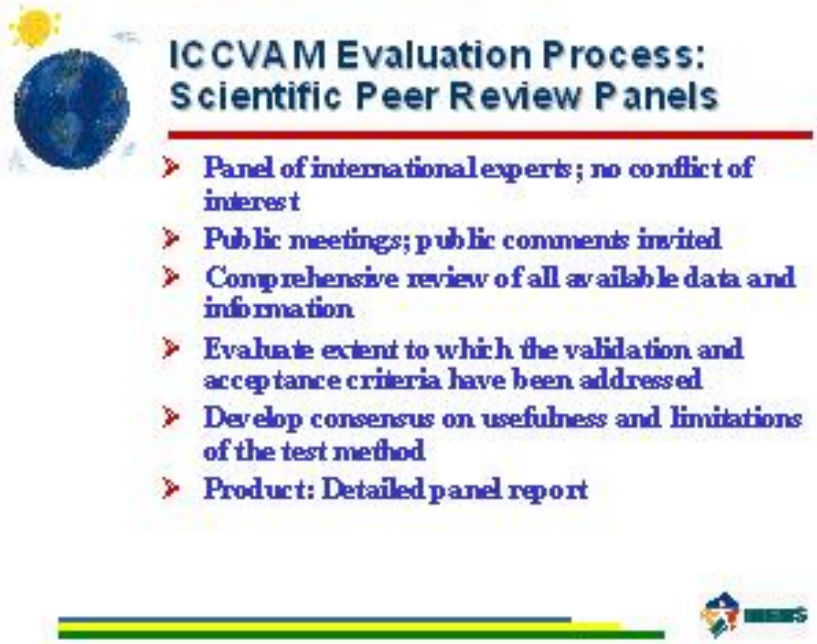


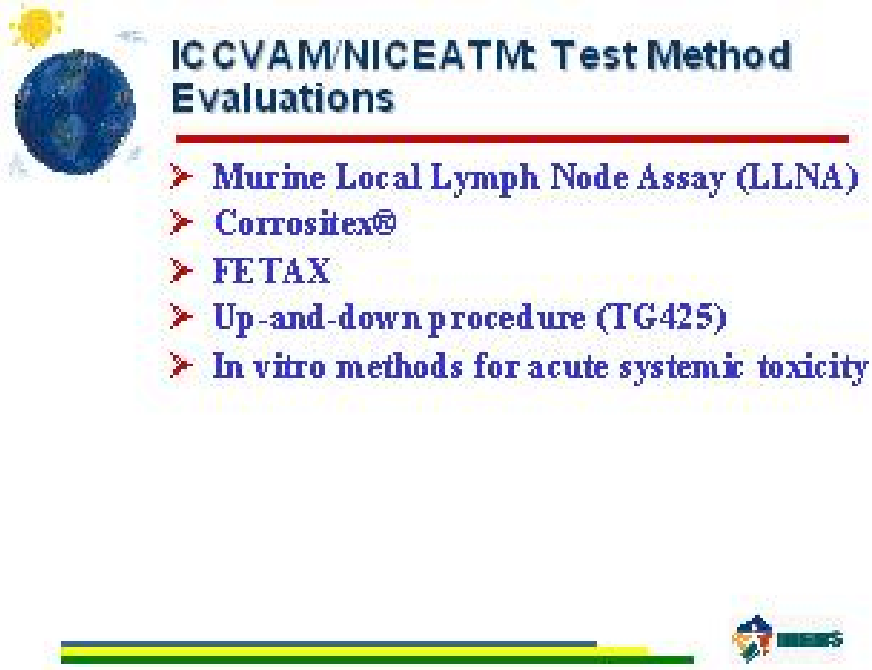


ICCVAM Test Method Submission Guidelines[1,2]

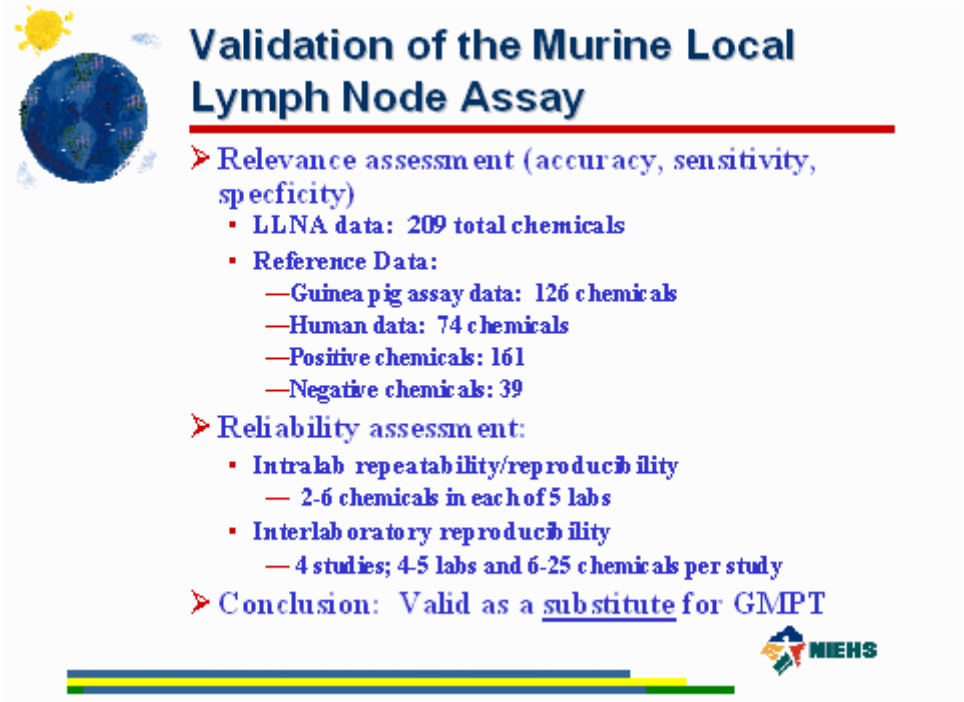
- Outlines format for test method background review document
- Background review documents
  - Provide data and information needed to assess a test method's current validation status
  - Provide basis for decisions on standardized protocols and validation study designs
  - Facilitate efficient and effective review!
- http://iccvam.niehs.nih.gov/docs/guidelines/subguide.doc

[1]Evaluation of the Validation Status of Alternative Toxicological Methods: Guidelines for Submission to ICCVAM, NIEHS, 1999; http://iccvam.niehs.nih.gov/docs/guidelines/subguide.pdf
[2]Federal Register: December 2, 1999 (Volume 64, Number 231, Page 67570-67580) [wais.access.gpo.gov]


ICCVAM Evaluation Process:
Scientific Peer Review Panels
Panel of international experts; no conflict of interest
Public meetings; public comments invited
Comprehensive review of all available data and information
Evaluate extent to which the validation and acceptance criteria have been addressed
Develop consensus on usefulness and limitations of the test method
Product: Detailed panel report


ICCVAM/NICEATM Test Method Evaluations
Murine Local Lymph Node Assay (LLNA)
Corrositex®
FETAX
Up-and-down procedure (TG425)
In vitro methods for acute systemic toxicity


The Murine Local Lymph Node Assay:
The Murine Local Lymph Node Assay
Sponsors
Dr. F. Gerberick, P&G
Dr. D. Basketter, Unilever
Dr. I. Kimber, Zeneca
Peer review panel meeting
September, 1998
Final Report: Feb. 1999
Regulatory acceptance
U.S.: October, 1999
OECD TG; June 2001
Implementation workshop
January 25-26, 2001
http://iccvam.niehs.nih.gov/methods/lnadocs/lnarep.pdf
NIEHS


Validation of the Murine Local Lymph Node Assay
Relevance assessment (accuracy, sensitivity, specficity)
LLNA data: 209 total chemicals
Reference Data:
—Guinea pig assay data: 126 chemicals
—Human data: 74 chemicals
—Positive chemicals: 161
—Negative chemicals: 39
Reliability assessment:
Intralab repeatability/reproducibility
— 2-6 chemicals in each of 5 labs
Interlaboratory reproducibility
— 4 studies; 4-5 labs and 6-25 chemicals per study
Conclusion: Valid as a substitute for GMPT
NIEHS

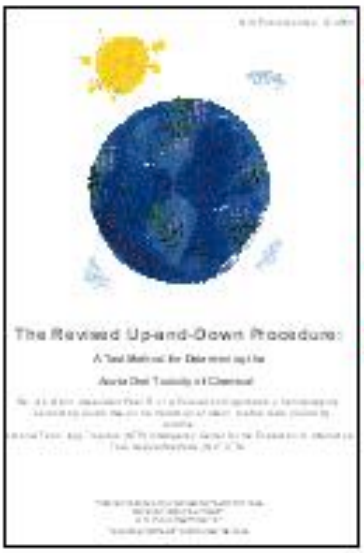
Corrositex®

- Sponsor
  - In Vitro International, Inc.
- Peer Review Panel Meeting
  - January, 1999
  - Final Report: June 1999
- Regulatory acceptance
  - CPSC, OSHA, Oct. 99
  - DOT, renewal, 1999-2000
  - OECD TG in preparation
- http://iccvam.niehs.nih.gov/docs/reports/corprrep.pdf

Validation of Corrositex®: An In Vitro Test Method for Dermal Corrosivity

- Relevance assessment
  - Corrositex® data: 419 chemicals
    - —IVI studies, ECVAM studies
  - Reference data:
    - —rabbit corrosivity tests: 163 chemicals
    - —positives: 89
    - —negatives: 74
- Reliability assessment
  - 3 labs, 60 chemicals

NIEHS

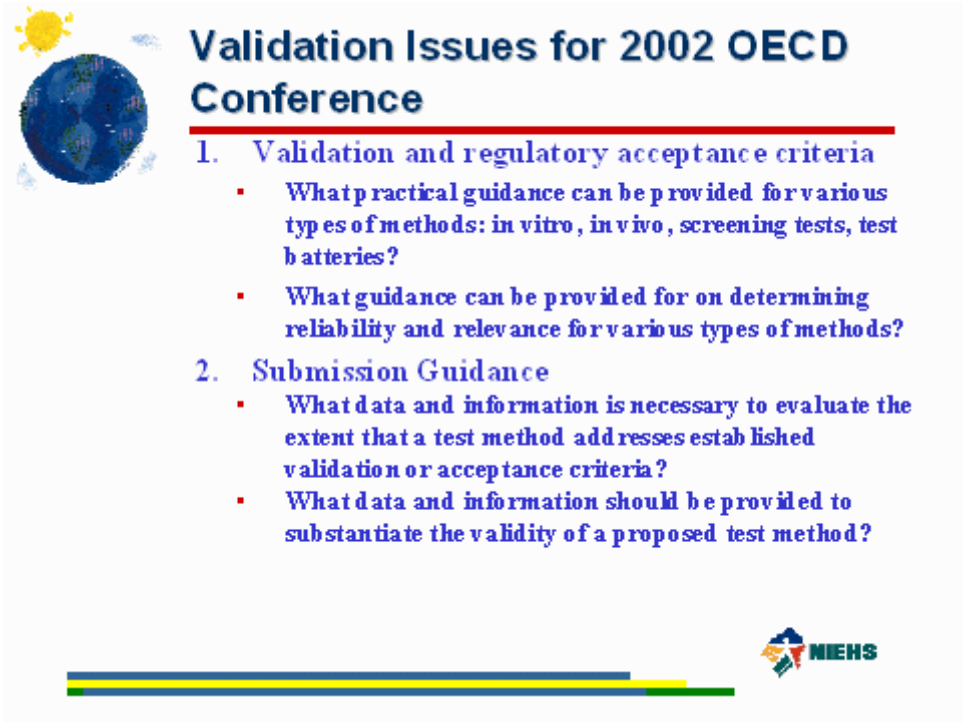


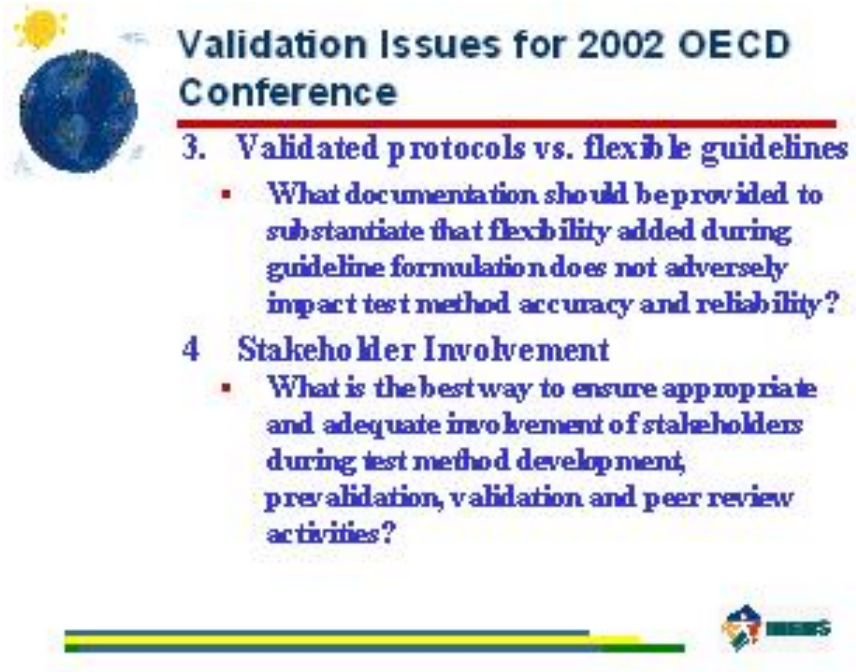
The Revised Up-and-Down Procedure

- Revision of TG425
- Peer Review Panel Meetings
  - July, 2000
  - August, 2001
- Training Workshop
  - February 19-21, 2002
- http://iccvam.niehs.nih.gov/methods/udpdocs/udpfin

Validation of the Revised Up-and-Down Procedure

- Based on computer simulations for new prediction models
- No animals were used!


Validation Issues for 2002 OECD Conference
1. Validation and regulatory acceptance criteria
What practical guidance can be provided for various types of methods: in vitro, in vivo, screening tests, test batteries?
What guidance can be provided for on determining reliability and relevance for various types of methods?
2. Submission Guidance
What data and information is necessary to evaluate the extent that a test method addresses established validation or acceptance criteria?
What data and information should be provided to substantiate the validity of a proposed test method?
NIEHS


Validation Issues for 2002 OECD Conference
3. Validated protocols vs. flexible guidelines
What documentation should be provided to substantiate that flexibility added during guideline formulation does not adversely impact test method accuracy and reliability?
4 Stakeholder Involvement
What is the best way to ensure appropriate and adequate involvement of stakeholders during test method development, prevalidation, validation and peer review activities?

Validation Issues for 2002 OECD Conference

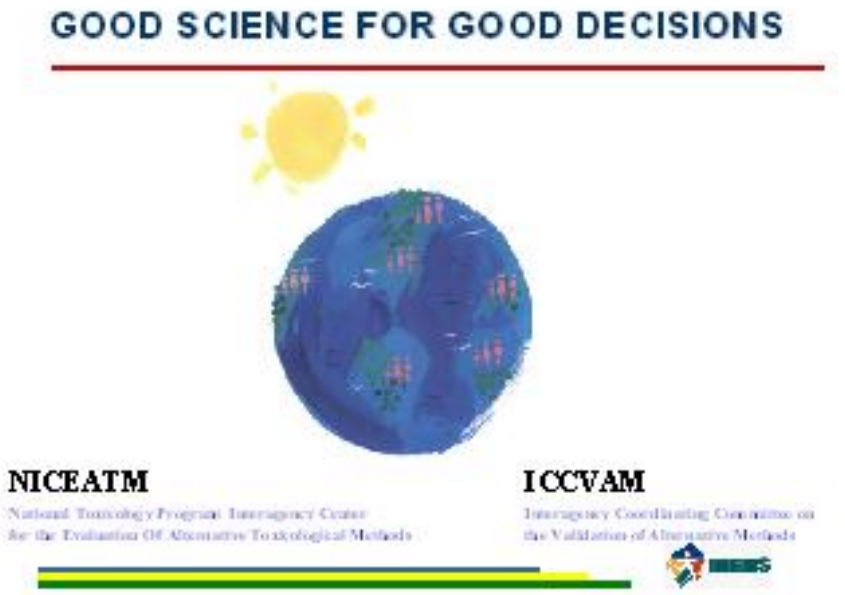
5. Patented test methods
   - Should policies be amended to allow for their inclusion in international test guidelines?
6. Guidance for adequate independent scientific peer review
   - Provision of test method submission to stakeholders
   - Opportunity for stakeholder input
   - Selection of experts without conflict of interest
   - Detailed documentation of reviews

⇒ What practical advice can be provided based on our collective experiences?

What technologies may help improve testing methods?

- Molecular biomarkers
  - Toxicogenomics
  - Proteomics
  - Metabonomics
- Transgenic models
- High throughput technologies
- Noninvasive imaging/labeling techniques
- Tissue engineering
- Bio-informatics
- QSAR


Why should we seek improved test methods?
Improved hazard/safety assessments
Reduced uncertainties in risk assessments
Increased prevention of injury and disease
Improved efficiency
Cost and time
Improved animal welfare
Refinement: less pain and distress
Reduction and replacement


GOOD SCIENCE FOR GOOD DECISIONS
NICEATM
ICCVAM

ANNEX 10

Mark Chamberlain, Unilever, UK

CHALLENGES AND ISSUES RELEVANT TO INDUSTRY

MARK CHAMBERLAIN and HARRIET WALLACE

Risk Analysis Group
Safety and Environmental Assurance Centre
Unilever Colworth
Sharnbrook
Bedford
UK

SEAC

Acknowledgements

Thanks are extended to Dr Julia Fentem, SEAC and Huner Gulay, Corporate Relations Department, Unilever for helpful advice and comment in developing this paper.

SEAC

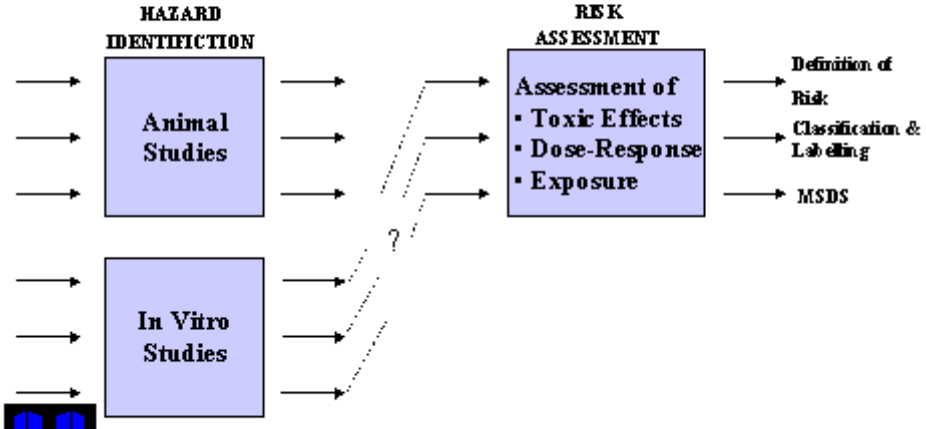
What has changed since the 1996 Solna Workshop?

Prediction Models

- been accepted generally
- now seen as an essential component of alternative test methods (Worth & Balls 2001)

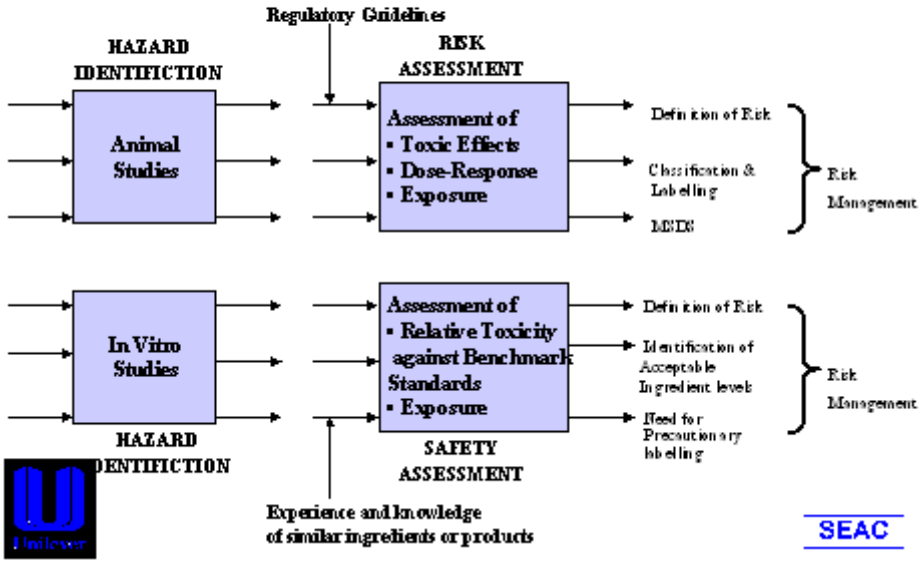
SEAC

A discontinuity in the hazard identification/risk assessment process



SEAC

The role of alternative test methods in safety assessment



What has changed since the 1996 Solna Workshop?

Several alternative test methods have been validated, scrutinised and accepted for use

SEAC

Alternative test methods have been validated, scrutinised and accepted for use for:

- Low molecular weight allergens
  - Local lymph node assay
- Phototoxic potential of chemicals
  - 3T3 neutral red uptake

SEAC

and for:

- Skin corrosive chemicals
  - Rat skin TER assay
  - Human skin models (EPISKIN, EpiDerm)
  - Corrositex (for acids, bases and derivatives)

SEAC

What has changed since the 1996 Solna Workshop?

In 1998, OECD consultation with selected experts as a follow up to the 1996 Solna Workshop

SEAC

Output:

"Draft Guidance document on the development, validation and regulatory acceptance of new and updated internationally acceptable test methods in hazard assessment"

OECD Environment, Health & Safety
Publications Series on testing and Assessment
No. 34

SEAC

Recent History

EARLY 1980's

- Some academic and industry interest
  - eye and skin irritation
  - LD50
- Some public awareness of animal testing

LATE 1980's

- Increased academic, industry and regulatory interest
- Vociferous anti-animal testing lobby and pressure groups emerged

SEAC

Recent History... cont'd

EARLY 1990's

- Industry started major investment in development and validation of alternative test methods
- Specific campaigns mounted by pressure groups

MID 1990's

- Industry Trade Associations driving hard with validation studies
- Active dialogues between politicians, regulatory, industry, academics and pressure groups

SEAC

Recent History… cont'd

2000
Selected test methods accepted as validated for use.

SEAC

Timelines for test method development, validation and acceptance

1. Local lymph node assay around 15 years
2. TER for skin corrosivity around 15 years
3. 3T3 NRU assay for phototoxic potential around 10 years (but based on earlier work over preceding 10 year on cytotoxicity tests).

SEAC

A challenge for Industry

With "15 year" timelines, the challenge is to be pro-active.

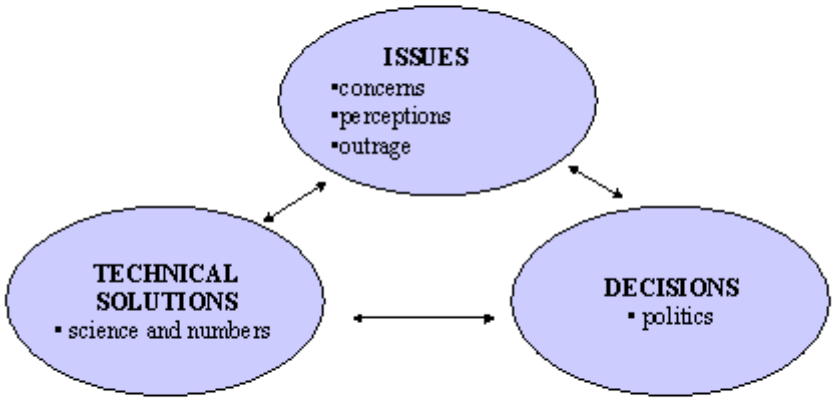
But what does this mean?

What is the overall context?

If scientists are not pro-active, problems can arise.

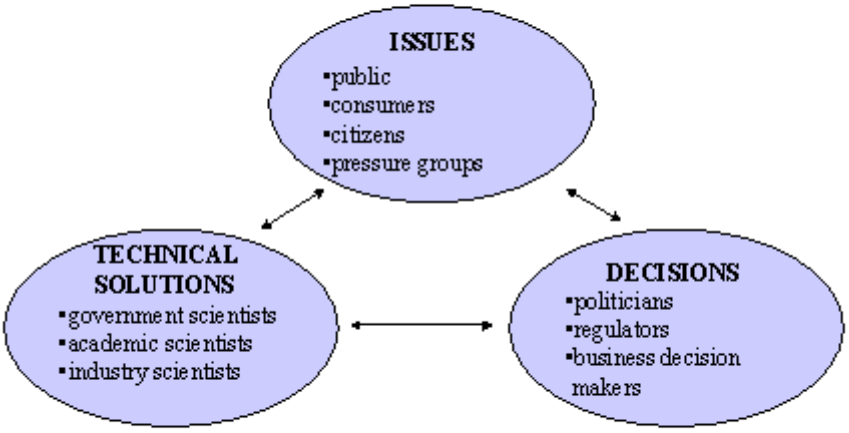
SEAC

Risk Communication Framework (What)



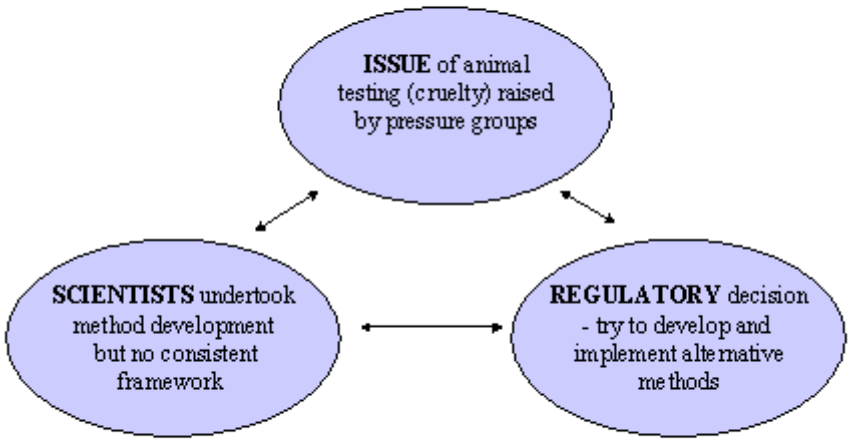
SEAC

Risk Communication Framework (Who)



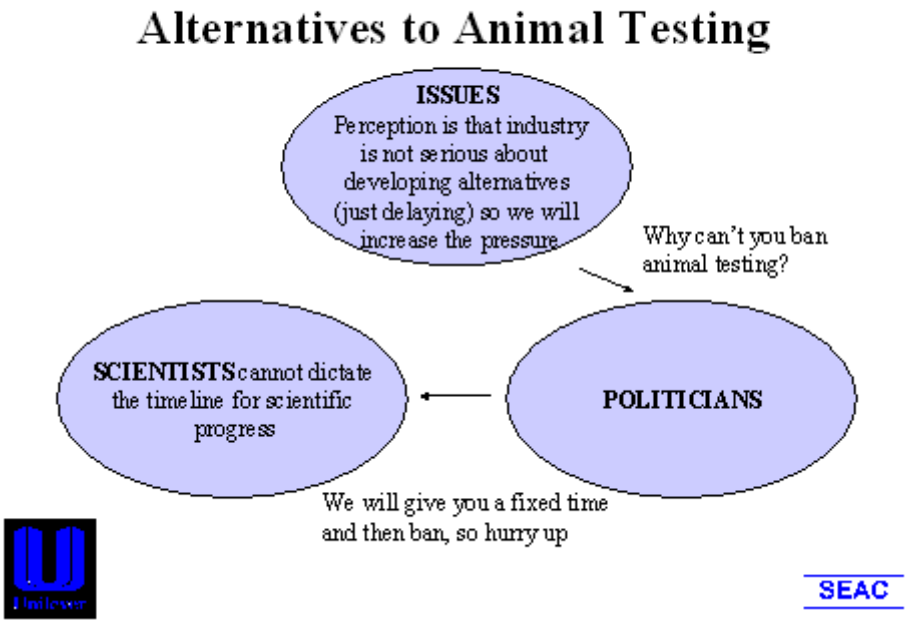
SEAC

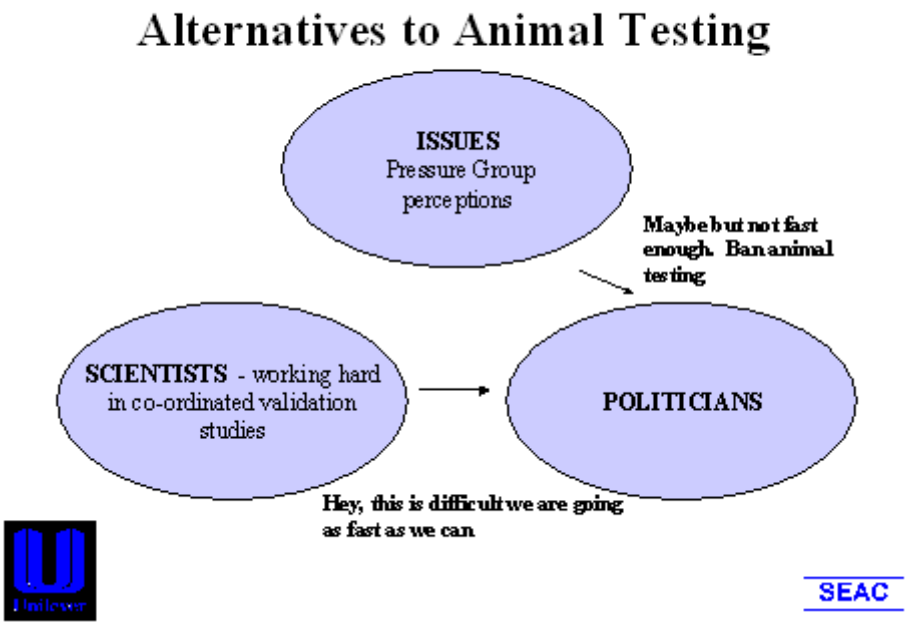
Alternatives to Animal Testing

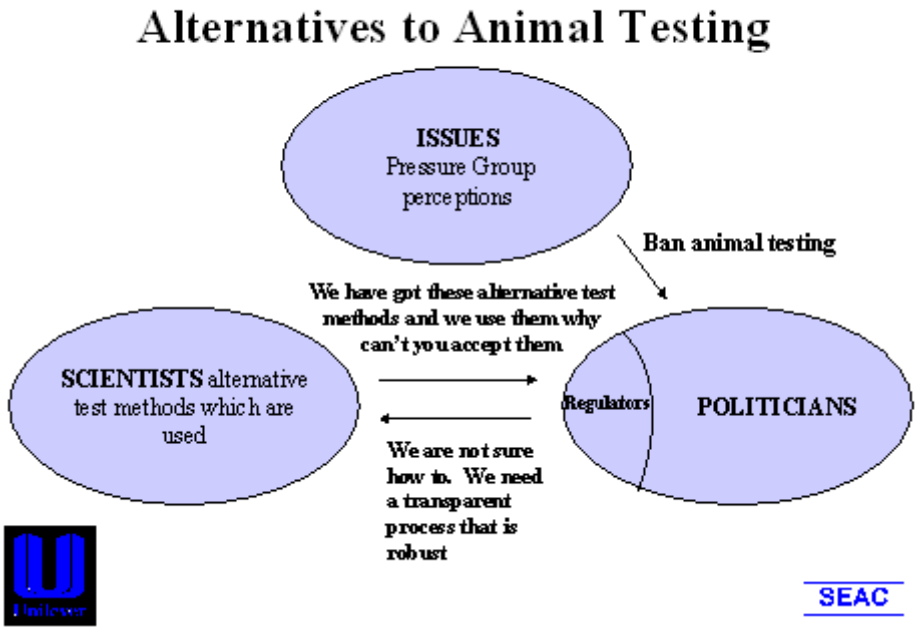


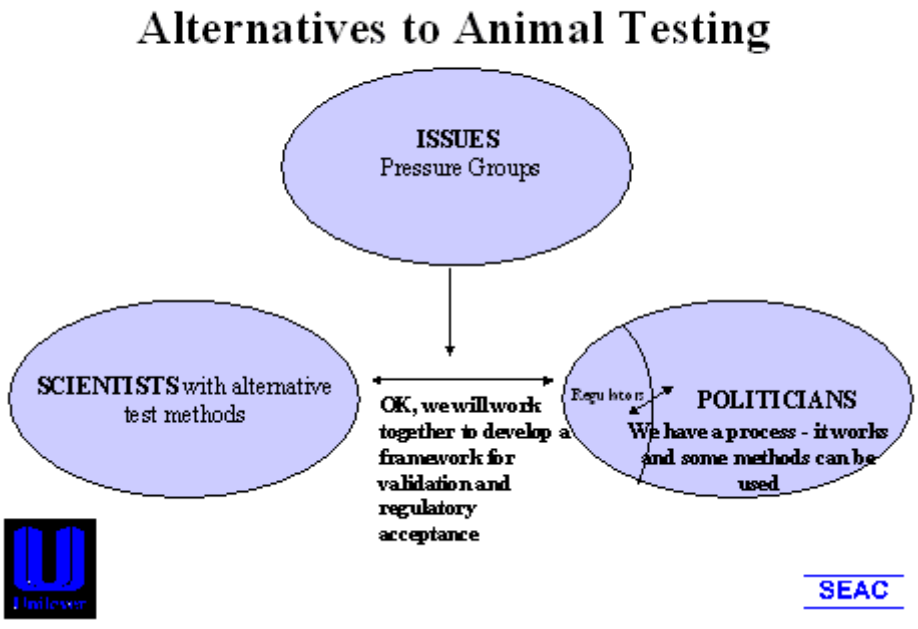
Solution: Co-ordinate validation studies BUT it did not happen fast enough

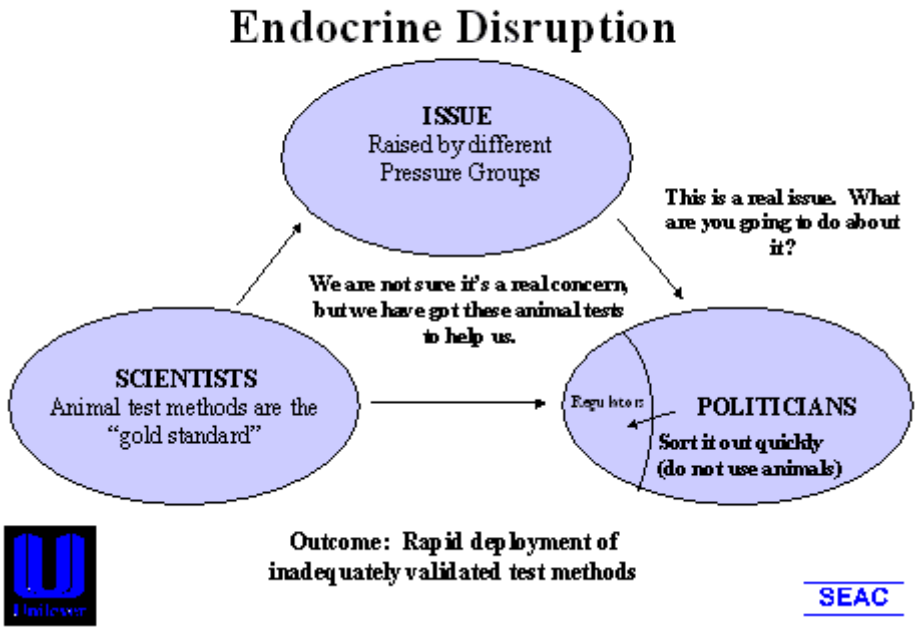
SEAC

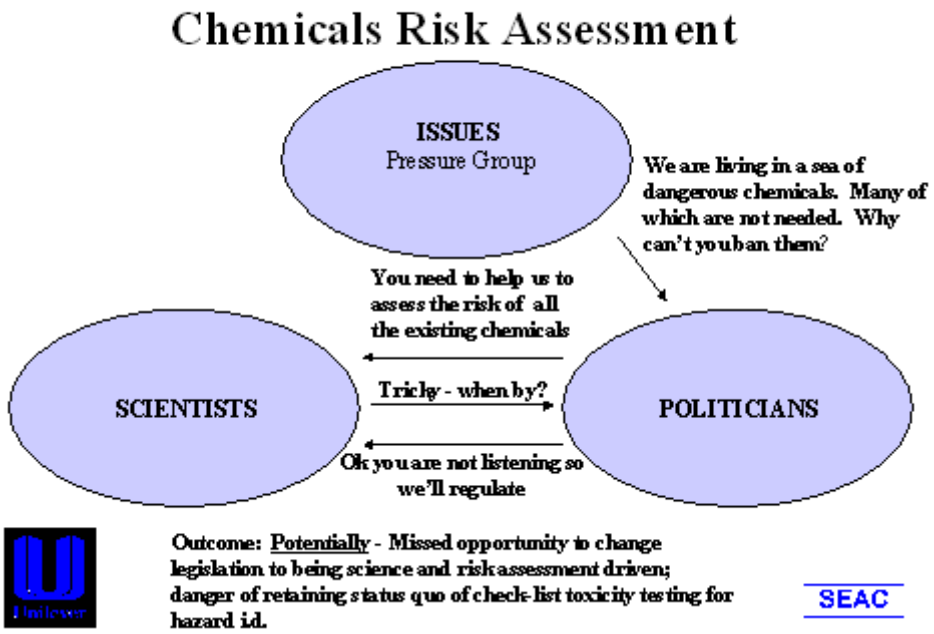

Alternatives to Animal Testing
ISSUES
Perception is that industry is not serious about developing alternatives (just delaying) so we will increase the pressure
Why can't you ban animal testing?
SCIENTISTS cannot dictate the timeline for scientific progress
POLITICIANS
We will give you a fixed time and then ban, so hurry up
SEAC


Alternatives to Animal Testing
ISSUES
Pressure Group perceptions
Maybe but not fast enough. Ban animal testing
SCIENTISTS - working hard in co-ordinated validation studies
POLITICIANS
Hey, this is difficult we are going as fast as we can
SEAC

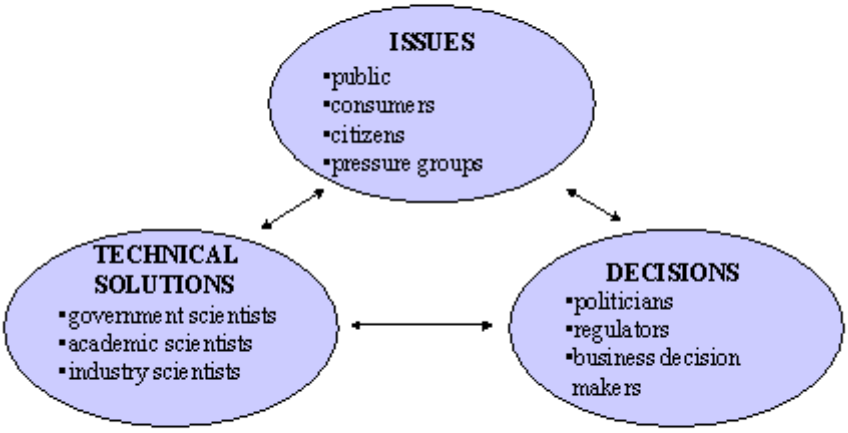

Alternatives to Animal Testing
ISSUES
Pressure Group
perceptions
Ban animal testing
We have got these alternative test methods and we use them why can't you accept them
SCIENTISTS alternative test methods which are used
Regulators
POLITICIANS
We are not sure how to. We need a transparent process that is robust
SEAC


Alternatives to Animal Testing
ISSUES
Pressure Groups
SCIENTISTS with alternative test methods
OK, we will work together to develop a framework for validation and regulatory acceptance
Regulators
POLITICIANS
We have a process - it works and some methods can be used
SEAC


Endocrine Disruption
ISSUE
Raised by different
Pressure Groups
This is a real issue. What are you going to do about it?
We are not sure it's a real concern, but we have got these animal tests to help us.
SCIENTISTS
Animal test methods are the "gold standard"
POLITICIANS
Sort it out quickly (do not use animals)
Outcome: Rapid deployment of inadequately validated test methods
SEAC


Chemicals Risk Assessment
ISSUES
Pressure Group
We are living in a sea of dangerous chemicals. Many of which are not needed. Why can't you ban them?
You need to help us to assess the risk of all the existing chemicals
Tricky - when by?
Ok you are not listening so we'll regulate
SCIENTISTS
POLITICIANS
Outcome: Potentially - Missed opportunity to change legislation to being science and risk assessment driven; danger of retaining status quo of check-list toxicity testing for hazard id.
SEAC

Risk Communication Framework



SEAC

Overall Challenge for Industry

For Chemicals

- Identify the hazard(s)
- Targeted risk assessment needs to be science based.

SEAC

Specific Challenges for Industry

For Risk Assessment

- Clear, transparent and unambiguous definition of risk assessment algorithms
- Acceptance of relative risk assessments using benchmark standards

SEAC

Specific Challenges for Industry... cont'd

- Publication of how data from alternative test methods are used for risk assessment (Jones et al, 2001)
  - Would seem to be relatively few risk assessment algorithms for use of animal data
  - HERA has demonstrated value of default values for exposure assessments

SEAC

In Conclusion

What does a WIN look like from this conference:

Use the experiences and learnings from the 1996 Solna Workshop and the contents of the OECD draft document - series on Testing and Assessment No. 34 - as the accepted framework for development, validation and regulatory acceptance of alternative test methods.

SEAC

In Conclusion

Why?

Industry needs agreement on this framework document for how alternative test methods are developed, validated and accepted by regulatory agencies.

So that the legitimate needs of the political domain and the issues domain (pressure groups) can be met.

SEAC

Outcome

If we reach this agreement then we can concentrate on the next real challenge - that of sorting out the risk assessment processes which is where we should now be focussing our efforts.

SEAC

ANNEX 11

SUMMARY REPORTS AND/OR SUMMARY STATEMENTS OF THE BREAKOUT GROUPS
BREAKOUT GROUP 1: PRINCIPLES AND CRITERIA FOR NEW AND UPDATED TEST METHODS

Co-Chairs: Dr Phil Botham (Zeneca CTL, UK); Dr Dave Hattan (FDA, US)
Co-Rapporteurs: Dr Rodger Curren (IIVS, US); Dr Gilly Griffin (CCAC, Canada)

1. The workshop participants in the Breakout Group 1 discussions included: B. Özturk Kyraci, Calvin Willhite, Dincer Karavut, Ibrahim Chahoud, Ih Chu, Jan van der Valk, Jun Kanno, Karl Jensen, Lars Wårngärd, Leon Bruner, Mark Chamberlain, Neil Wilcox, Susan Hazen, Troy Seidle, Ursula Sauer, Mark Jaber, Jean Roch Meunier, Taisen Iguchi, and Eric Vindimian.

2. This listing may not be accurate or complete because a number of individuals initially assigned to this Group opted to participate in other groups, or shared their time among Groups. Similarly, some individuals who were not originally assigned to this group participated in this Group's discussions.

Summary Report/statements

3. The following summaries the discussions and recommendations of Breakout Group 1. The questions supplied to the Group, and questions and comments from the various member countries, were discussed and considered.

Validation

4. All types of tests need validation. These tests include:

- Human tests
- In vivo tests
- New tests
- Substitute tests
- Modification of endpoints
- Multiple endpoints
- SAR
- Genomics/proteomics, etc.
- Statistical methods
- Test batteries
- Tiered tests
- Ecosystem (community level)

5. Most, but not all, of the tests on the following list were reviewed by the Group, and a number of validation principles were affirmed or clarified, including:

- The 'Test development' section should appropriately be labeled 'prevalidation'. A section on prevalidation is missing; this is needed because it sets the stage for the overall document.
- Guidance should be provided for chemical selection in prevalidation and validation stages.
- It should be noted that test validation is independent of risk assessment.
- Flexibility is not synonymous with ambiguity. There should be flexibility in the approach, not in the principles.
- Relevance includes knowing the uncertainties involved in the test. Suspect relevance is not sufficient; there should be empirical proof that the test is relevant. Relevance is inherent in large observational studies because the apical effects are measured.

- It was recommended that there also be peer review of protocols prior to starting the validation study, and that protocol changes during the study should be avoided.
- The same validation standards should be used for *in vitro* and *in vivo* tests.

6. There is no one proper authority to do peer review; it can be performed by any competent group. The independent peer review makes statements on the validity of the method. The peer review panel reports to the VMG, and the report is publicly available. The process should be transparent regardless of who performs the peer review.

7. There was an extensive discussion of purpose of validation, and the Group reaffirmed that validation should be a hypothesis-driven process that utilizes the "Solna" criteria. A flow chart (Figure) was prepared that illustrates the validation process and the inter-relationships of the various validation phases and steps. The Group recommended that the new Figure replace the current Figure 1 in the draft Guidance Document, and that the associated paragraphs be revised.

8. It should be noted that one can cycle through the validation loop, as described in the Figure, without further laboratory experiments, but this could lead to over-learning. Cycling back through the scheme too often can create a self-fulfilling prophecy, in that the prediction models may be forced to fit the current set of test chemicals. This could result in an assay that is not generally useful. This caution should be added to the document.

9. The Guidance Document should address the situation of the validation of a new test when no appropriate reference data are available. One recommendation was to use related (or complementary) endpoints. A more complete explanation is needed of the terms used in the document.

The prediction model[1]

10. The Group extensively discussed the construction of a prediction model, and how to test the hypothesis of whether it works. One of the Solna criteria is that the relationship of the test method's endpoint(s) to the biological effect and toxicity of interest must be addressed. A prediction model describes the relationship between the results of the test and some toxicological concern. It defines the steps to be taken to convert results from a toxicity test method into a prediction of toxicity useful for making decisions. This criterion supports the use of a prediction model which could be qualitative or quantitative.

11. As a result of extensive discussions regarding the concept of the prediction model, and as to whether the term 'prediction model' was well understood and applied to all types of tests, the following recommendations were made to the Plenary session:

- There is a need to give guidance on how to (or whether it is possible to) construct prediction models for new assays that do not have available reference data, e.g., developmental neurotoxicity.
- There should be a process to achieve wider understanding of the concept and utility of the prediction model. A proposal was made to organise a workshop or symposium at an international meeting.
- There was extensive discussion about whether the term prediction model clarifies or muddles the discussion. New terms should be introduced only when needed, and the point was made that, the term retards rather than aids. The prediction model is, essentially, a statement of a

1. See Figure 1: Proposal for Breakout Group 1 for a Flowchart

hypothesis, and it is important to separate the concept from the terminology. As a result, an alternative and more widely understood descriptive term for the prediction model concept should be adopted.

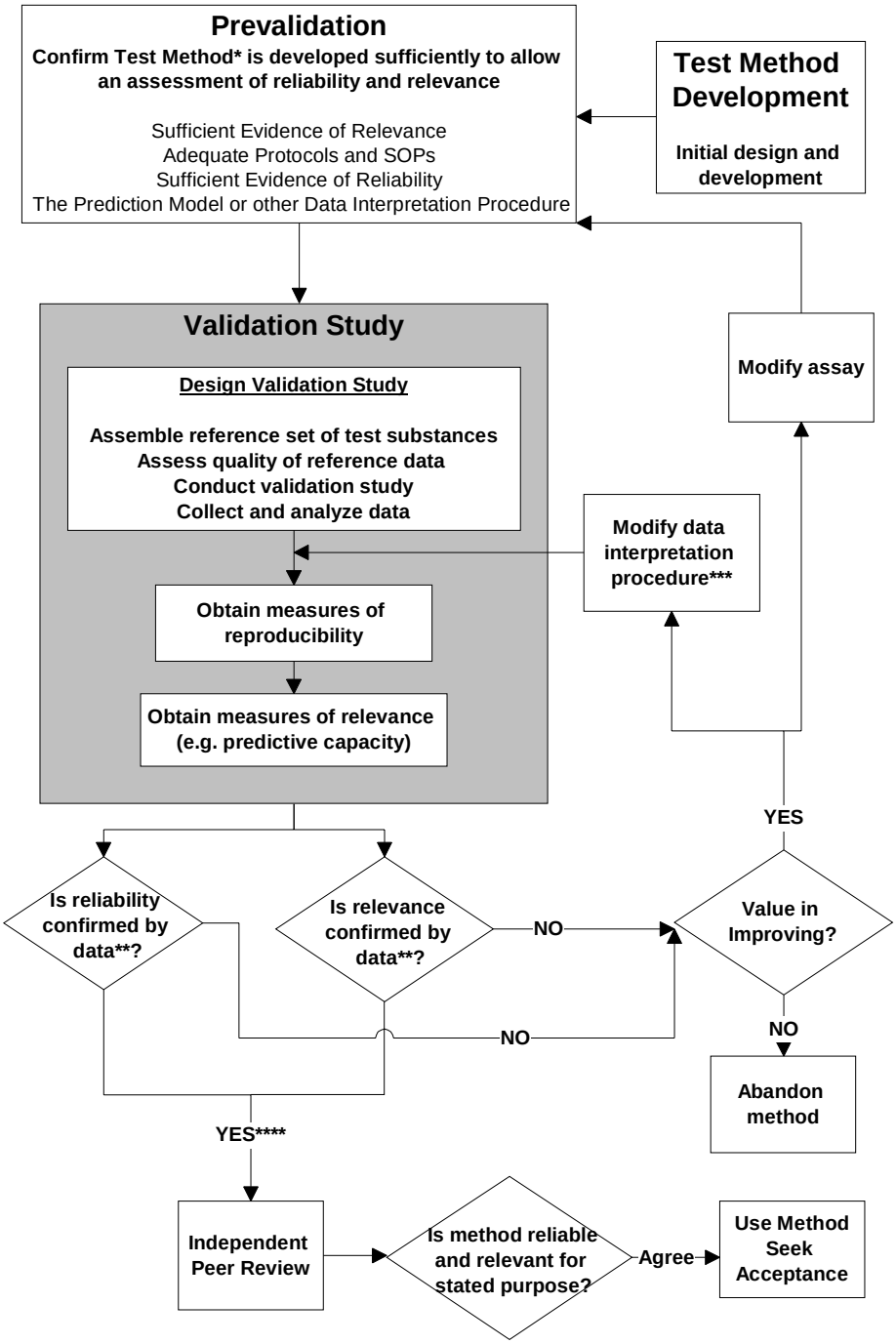
- Examples of qualitative and quantitative prediction models should be presented.

12. Examples of prediction models provided by the Group for incorporation into the Guidance Document include:- A statistically significant increase in uterine weight is indicative of estrogenic activity.

13. Based on the LLNA validation: When the scintillation count of the treated group is three fold or greater than the scintillation count of the negative control, the substance is a dose skin sensitizer. - Based on a Carcinogenicity Assessment Study: A tumorigenic response is indicated when the tumor occurrence rate of a treated group is statistically greater ($p<0.05$) than the same response in the control group.

14. General studies: The lowest dose at which a toxicological endpoint is determined (based on a statistically significant difference between the control and a treated dose response) is the lowest observed effect level (LOEL).

Figure 1.
Proposal from Breakout Group 1 for a Flowchart
(although discussed in plenary, time did not allow to reach a consensus of the flowchart)



*Test method is the assay and Prediction Model as described in Worth and Balls, 2001. *
*The confirmation of reliability and relevance can occur in any order or in parallel. This assessment is conducted by the validation management group. If validation study confirms reliability and relevance the results of the study can be evaluated in an independent peer review.
***Caution: Overfitting of data to a model may diminish the general application of test method
****The data need to confirm reliability and relevance in order to proceed to independent review

BREAKOUT GROUP 2: PRACTICAL GUIDANCE ON THE MANAGEMENT AND CONDUCT OF THE VALIDATION PROCESS

Co-Chairs: Dr Julia Fentem (Unilever, UK); Dr Willie Owens (Procter & Gamble, US)
Co-Rapporteurs: Dr Bob Combes (FRAME, UK); Dr Rochelle Tyl (RTI, US)

1. The workshop participants in the Breakout Group 2 discussions included: Don MacGregor. Erik Walum, Eva Sandberg, Marike Kolossa, Martin Stephens, Roger McClellan, Ronald Joiner, Sylvie Tissot, Tommy Stagh, and Tony Maciorowski.

2. This listing may not be accurate or complete because a number of individuals initially assigned to this Group opted to participate in other groups, or shared their time among the different Groups. Similarly, some individuals who were not originally assigned to this group participated in this Group's discussions.

3. The following topics and questions were developed by the Steering Committee to guide the Breakout Group's discussions. They were made available to all participants in advance of the Workshop.

Summary/Report Statements

4. The following summarises the discussions and recommendations of Breakout Group 2.

General comments

5. The introductory paragraphs should address the organisation of validation activities, which can range from small-scale (i.e., individuals or groups of investigators) to large-scale, multi-national groups. The purpose of the Guidance Document is to provide a broad, generic document which also provides examples.

6. The document should reflect that the primary purpose of a test is not replacement, etc. It is the assessment of human or ecological hazard.

7. A new bridging section is needed at about paragraph 27 to indicate that the overall principles of the validation of regulatory tests are applicable to different bodies and groups.

8. The use and purpose of a test, as well as the endpoints selected, must be clear in order to plan a validation study, e.g., to review the most suitable information and data, and to develop an appropriate prediction model.

9. The definition of a test must be broadened to cover all types of tests, and should not be restricted to mechanistically-based ones. Some of the basic terminology (e.g., definition of a test) needs re-wording so as to help those completely new to validation.

10. All OECD Test Guidelines should be based on validated test methods. There needs to be a process to facilitate the incorporation of a newly validated test into a regulatory guideline. The Guidance Document should have an Annex that outlines relevant OECD processes using illustrative flow charts, etc., and describing the roles of the National Co-ordinators.

The validation process

11. The Figures in the draft document should be replaced with flow charts and terminology from the Solna document to the extent possible, but there should be less detail and more continuity shown in the process, with careful use of terminology. The following should be shown:

- sequence of events and decision points in the overall process of a validation study from test development to completion of formal validation;
- stages of peer-consultation and peer-review;
- references to case-studies and other background documents in the Annexes;
- references to relevant consecutive paragraphs in the text;
- places where flexibility in approach can be applied;
- different entry points into the process for tests in different stages of prior development, and for different uses of a test.

12. The validation process can be used retrospectively for tests that have been accepted by convention but have not been previously validated, or for tests that have never been formally validated.

13. The Guidance Document should be provide information on validating test batteries. If test batteries are being proposed, this must be made clear at the start of a validation study, and the individual component tests should be validated before the validation of the battery can be addressed.

14. It is strongly recommended that validated tests be reviewed periodically as new information becomes available.

Validation management group, peer review, and involvement of interested parties

15. The Guidance Document should emphasise that: (a) a VMG would normally be comprised of members who are involved in the development of the test method(s) being evaluated in a validation study; (b) peer consultation can be conducted periodically during the conduct of a study; and (c) peer review should be a totally separate scientific process that should be conducted after the completion of the study by a different body, whose members are largely independent of the study (e.g., as conducted by ECVAM in Europe and ICCVAM in the US). The Guidance Document also should make explicit that all the comments in the peer review report should be addressed, and that both the report and the response should be in the public domain.

16. The planning of a validation study should be undertaken on a case-by-case basis to determine the individual components of validation that need to be included (e.g., the entry and exit points in the overall scheme of validation), according to the nature of the test, its intended usage, and the extent of its prior validation.

17. Involvement in test development and validation should be open and flexible. All intended major end-users of a test (e.g., regulators) should be involved as early as possible in the design of the validation study to facilitate the eventual acceptance of the test method after it has been validated.

18. Appropriately qualified groups should be involved in the planning and conduct of validation studies, irrespective of whether or not they are National centres for validation. The need for and role of a validation management group will depend on the scope of the validation study and the purpose of the test.

19. Regulatory authorities should be consulted in the planning of a validation study. Regulatory acceptance follows peer review, but not automatically. Therefore, consultation with regulatory authorities is expected to facilitate eventual acceptance of the test.

20. More clarity is needed on the link between the peer review and the regulatory acceptance processes (especially the mechanism whereby a new Test Guideline can be drafted and submitted for approval and acceptance). Regulatory agencies have diverse procedures for converting validated, peer-reviewed test protocols into regulatory Guidelines; this should be acknowledged and addressed in the document.

21. Peer review is termed 'largely independent' to allow input into the process from people or groups who may not be independent.

22. The Group agreed that there should be the opportunity for public comments to be submitted with regard to newly-planned validation studies. Public notification of planned validation studies should be encouraged in order to get appropriate stakeholder involvement prior to the commencement of the studies. Cultural, political, and legal differences among Member Countries need to be taken into consideration.

23. The validation study report should be published in the peer-reviewed scientific literature following the normal manuscript review process. Publication of the results of the validation study is a separate process from peer-review for validation and public availability of the validation study.

Reference chemicals

24. A validation study is generally designed to demonstrate that a test method can accurately predict the activity of chemicals (i.e., test relevance). The main criterion for selecting chemicals should be the need to demonstrate that the assay works for its intended purpose; this has to be done on a case-by-case basis. Therefore, the chemicals used in the validation study should be appropriate for the endpoint(s) being measured by the test. The Guidance Document should also take account of the need to evaluate mixtures of chemical substances.

25. The wording of paragraphs 56-64 is satisfactory.

26. The wording of paragraph 60 needs to address the possibility that reference chemicals with a full range of activities and potencies may not be available. If the desired chemicals are not available, statistical advice should be used to ensure that the objectives of the study can be met with the numbers and types chemicals available. Alternatively, it may be possible to use reference chemicals with data derived from related endpoints.

27. If the test chemicals are coded, there must be sufficient information provided for each so that they can be tested properly, and that the safety of the laboratory staff can be adequately protected, taking into account local and national regulations. Thus, all relevant physico-chemical and safety information should be made available to all participating laboratories.

28. The Group noted that current guidance in paragraphs 36-37 regarding reference data is not sufficient, but did not offer alternative text.

Statistical support

29. The section of the draft document on statistical guidance and support (paragraphs 33, 34) is confusing. The VMG should involve an appropriately qualified statistician in the study design, development of performance criteria, and selection and implementation of statistical analytical procedures. This statistician does not have to be a formal member of the VMG.

30. The protocol should specify the statistical analytical methods to be used. There should be an independent statistical analysis of the overall data, which should be conducted according to GLP principles (e.g., via software validation). To support this effort, a data management system for data collection, analysis, transfer, storage, distribution, and accessibility should be specified as part of the project plan for the validation study.

31. The approach adopted to assure statistical support should not be prescriptive. The VMG should have access to statistical advice by any appropriate method (e.g., VMG member; external consultant; lead lab statistician).

32. Statistical aspects of prediction models were briefly discussed. Examples from existing validation experiences should be used as case studies.

Laboratory qualifications and capabilities

33. The test is being validated, not the laboratory. Therefore, paragraphs 61-64 should be focused more on the test than on lab performance. Sufficient information, guidance, training, and explanation should be provided so that a common protocol can be performed identically in all participating laboratories. Compliance with the common protocol should be checked throughout the conduct of the study by periodic visits to participating laboratories by members of the VMG.

34. The record of participating laboratories in complying with relevant animal welfare standards and legislation should be one of the criteria for their selection (see paragraph 51).

GLP compliance

35. There is inconsistency throughout the draft document for the need for GLP compliance (e.g., paragraphs 39 and 78). Consensus could not be reached by the Group on the need for GLP. Some Group members believed that all participating laboratories in a full validation study should be fully GLP compliant. Others believed that the participating laboratories should be GLP compliant as far as possible; where not, there should be documentation as to where compliance was not complete. However, the overall management of the study should follow GLP principles.

36. When planning a validation study, the VMG should make a decision on the GLP status of the participating laboratories, the decision should be justified, and this information should be transparent.

Financial

37. The term ‘sponsorship’ is vague and should be carefully addressed. Sponsors could be financial (e.g., test owner) and/or management sponsors.

38. The provision of financial resources for validation studies has not kept pace with the needs, and this problem has to be addressed. The Guidance Document must be more explicit about budgetary issues arising during a validation study and the need for transparency in the provision of funding. The expensive

nature of validation studies needs to be recognized in the document, and suggestions made for encouraging the sharing of costs and active collaboration, to avoid duplication of effort.

Editorial recommendations

39. The draft document needs serious editing and re-structuring to ensure continuity and smooth transitions from section to section. This task would be facilitated by including people on the editing group who were not involved in drafting the initial document.

40. Paragraphs 50 - 53 need to re-ordered to the sequence: 50, 52, 53, and 51.

41. Paragraph 90 should include (Q)SAR

BREAKOUT GROUP 3 : PRINCIPLES AND CRITERIA FOR REGULATORY ACCEPTANCE OF VALIDATED TEST METHODS, INCLUDING THE SUBMISSION OF INFORMATION TO SUPPORT THEIR VALIDITY

Co-Chairs: Dr. Leonard Schechtman (FDA, US); Dr. Otto Meyer (Division of General Toxicology, Denmark)
Co-Rapporteurs: Dr. Abby Jacobs (FDA, US); Dr. Manfred Liebsch (ZEBET, Germany)

1. The workshop participants in the Breakout Group 3 discussions included: Anna Tompa, Betty Hakkert, Bob Litelpo, Christiane Aveline, Edmund Plattner, Gill Langley, John McArdle, Karin Gabrielsen, M. Dunier-Thomann, Odile de Silva, Ole Ladefoged, Peter Evans, Atsuya Takagi, Walter Diembeck, Jim Freeman, Tim Springer, and Vanessa Vu.

2. This listing may not be accurate or complete because a number of individuals initially assigned to this Group opted to participate in other groups, or shared their time among Groups. Similarly, some individuals who were not originally assigned to this group participated in this Group's discussions.

Summary report/statements

3. The following topics and questions were developed by the Steering Committee to guide the Breakout Group's discussions. They were made available to all participants in advance of the Workshop.

Proceedings of Breakout Group 3

4. The following summarises the discussions and recommendations of Breakout Group 3.

General comments

5. It was recommended that an executive summary of the Solna principles be generated and appended to the Guidance Document, in recognition of the fact that the document is based on the Solna principles.

6. Section IV.4.a provides insufficient guidance on what constitutes appropriate flexibility. It is important to note that flexibility applies to the scope of the validation study, but not the scientific rigor. The draft Guidance Document (paragraphs 81, 82) should be revised to explicitly state that scientific rigor should always apply regardless of the scope of validation or the type of study (e. g., *in vitro* or *in vivo*, or whether the method is new or revised).

Validation

7. The scope of validation may vary (e.g., fewer chemicals may be needed for catch-up validation). The guidance provided in the draft document is better for full tests than for abbreviated tests.

8. The validation criteria for new tests were discussed. However, discussion is needed with regard to a test that has been extensively used by regulatory authorities but is now being modified or revised. There was a statement regarding the document term 'streamlined peer review'. This concept should be further defined and explained. Similarly, the concepts, new test and revised test should be defined and distinguished.

9. Paragraph 83 should be deleted because a “lower level of assurance” is not appropriate for any types of validation studies. Similarly, the standard for a retrospective validation (i.e., paragraph 89) should not be lowered. The group was of the opinion that the concept and the term be defined.

10. Independent peer review is an absolute requirement; the draft document presents it as an option. This must be changed.

11. A new section should be added after Section VI to provide guidance on the contents of a submission to support regulatory acceptance of a test method. It was also recommended that the ICCVAM submission guidelines document be appended as an example. The web address of the ICCVAM document should be cited. Headings in the table of contents of the ICCVAM submissions guidance document, p. i, III, 1.0-12.0, should be included as a table in the OECD document.

12. It was recommended that appropriate guidance regarding OECD Test Guidelines be provided if the Guidance Document is to designed to address those Guidelines, e.g ., the document should provide guidance on the transition from a specific, validated protocol to a more flexible OECD test guideline, and the document should address the possibility supplemental data may be needed to define and support changes to critical variables.

13. The document should provide guidance on the transition from a specific, validated protocol to a more flexible OECD test guideline. If necessary, as part of the validation process, an expert panel could be used determine how to allow the Test Guideline to be broader than the protocol that was validated.

14. Prior to initiating a validation study, OECD nominated experts should examine protocols for key elements and performance criteria

15. A statement should be included near the beginning of the document to address the need for early communication between test developer and regulatory authorities which could facilitate eventual acceptance of a test method.

Patented methods

16. Current OECD policy requires that the method must be freely available for use. The OECD restriction on developing Test Guidelines using patented methods include single source materials. However, if a patented method should be considered scientifically valid if it meets criteria for regulatory acceptance as outlined in Table 2 in the draft Guidance Document. If patented methods or materials are used in Test Guidelines, they should be readily available to potential users. The performance criteria of the patented method should be described and reference chemicals identified, so that a generic method could be developed, validated, and accepted.

17. The Group recommended that the current OECD policy regarding patented methods be reconsidered.

Recommended editorial revisions:

18. Table 2 in the draft Guidance Document is a key table and all points in it are applicable. Its title should be revised to "Principles and criteria for regulatory acceptance of a test method". Some clarification and additions to the table were suggested.

Clarifications

- The bolded text should be revised to read ".. subjected to a transparent independent peer review process." because transparency of the peer review process fosters greater acceptance.
- The definitions of "test" and "test method" are incomplete and were unclear to some Group members. It is recommended that the definitions of "test method" and "test" be modified to include the prediction model or the extrapolation process.
- A better definition of "prediction model" is needed.
- Bolded text should be unbolded because it does not add anything.

Additions

- Detailed protocols and SOPs should be available
- The purpose, strengths, and limitations of the test method should be specified and fully described.

19. The Group recommended that the title of the Guidance Document be changed, but a specific title was not recommended. However, the following modifications were suggested, "Draft guidance document on the development, validation, and acceptance of .....test methods for regulatory purposes...in hazard assessment."

20. Line 7-8 in paragraph 89 should be replaced with "For such a case, the assembled data should be evaluated according to the validation principles described above...."

21. Consistency of terminology throughout the document is needed, e.g., clearly distinguish between hazard and risk, revised and updated, new test, replacement test, and substitute test method.

22. Revise paragraph 104 for clarity, as there is confusion between national and international issues.

23. Paragraph 103 is redundant and should be deleted.

BREAKOUT GROUP 4: PRACTICAL GUIDANCE ON THE PROCESS FOR INDEPENDENT PEER REVIEW AND REGULATORY CONSIDERATION AND IMPLEMENTATION

Co-Chairs: Dr. Andreas Gies-Reuschel (Umweltbundesamt, Germany); Dr. Kathy Stitzel (Procter & Gamble, US)
Co-Rapporteurs: Dr. Karen Hamernik (EPA, US); Dr. Vera Rogiers (Vrije Universiteit Brussel, Belgium)

1. The workshop participants in the Breakout Group 4 discussions included: Brita Hagström, David Wilkins, Jennifer Seed, Kimmo Louekari, Lars Terenius, Lena Odland, Lorraine Twerdok, Michael Balls, Teiji Takei, Toini Berzins, Wally Hayes, Wolfgang Pape, and Yasuo Ohno.

2. This listing may not be accurate or complete because a number of individuals initially assigned to this Group opted to participate in other groups, or shared their time among Groups. Similarly, some individuals who were not originally assigned to this group participated in this Group's discussions.

Summary report/statements

3. The following summarises the discussions and recommendations of Breakout Group 4.

General comments

4. The draft Guidance Document is much too prescriptive and detailed, and should be simplified. As it is currently written, it does not support the concept of flexibility. There are may be ways to approach and perform these processes and the guidance should be more general, and not a check list of details. The important point are not who is doing what, but what needs to be done. There was full agreement that the basic validation principles (e.g., Solna; OECD; ECVAM; ICCVAM) are appropriate and workable. A simple Guidance Document should be written that describes those principles.

5. The organization of the document also needs to be considered. For instance the validation methods should be discussed together, e.g., the "validation through available data'' and "SAR" sections should be included in Section IV on validation.

6. The Group agreed that the GLP principles are applicable to validation studies.

7. The document should include a section regarding the deletion of obsolete OECD Test Guidelines.

8. The question of how to get widespread implementation of new methods is important and needs to be addressed, but is beyond the scope of this document.

9. The questions and issues regarding regulatory considerations were essentially political, and beyond the scope of the guidance document. Regulatory acceptance issues should also be separated from the peer review process.

The validation study and the validation management group

10. For clarification, the Group provided a Table (attached) describing the various stages of test validation as they are currently practiced in Europe, Japan, and the US.

11. Despite the fact that a regulatory authority may require a specific data set and has expressed confidence in a test method that provides that information, the need for an independent peer review process is not waived. The Group supported the need to validate all new or substantially modified animal and non-animal standard test methods before adoption.

12. The independent peer review and validation steps were muddied in the draft document and need to be separated. The management of a validation study does not to be independent, in contrast to paragraph 29 of the draft document; however, the peer review must be independent. The validation management committee may come from the industry or laboratory that developed the method, and thereby have a vested interest in the outcome of the study. As a result, the management group cannot be considered completely unbiased.

13. The test development, selection, and validation processes should be unbiased and transparent. It may not be reasonable or achievable to have all validation management group members be completely independent of the entire development and validation process. However, the interests of the participants must be clearly stated and transparent, because the panel as a whole must have credibility as unbiased scientific group.

14. Whether a protocol change is substantial enough to require a new validation study and peer review would depend on the significance of the change. If there is a question there should be consultation with a group of experts, including regulatory authorities, when appropriate.

15. There is more than one way by which tests can be validated, including the use of historical data and computer simulations. These possibilities should be added to the guidance document. The term "prediction model" may be too specific to apply to both *in vivo* and *in vitro* studies and needs to be more simple and flexible if it is going to apply to both. The use of the word 'algorithm' in the glossary implies a mathematical model which is not always necessary. The area of statistics is much too prescriptive, as currently written. Statistical expertise is necessary for a validation, but the draft document is much too specific on how this is to be brought into the management team and what the statistician should do.

16. Although it was recommended that the results of the validation study be published, peer review is sufficient for regulatory acceptance if the peer review report and the results of the validation study are publicly available. paragraph 102 should be rephrased as: "The final report of the peer review panel should be publicly available."

17. It was recommended that the results of validation studies be published in a peer review journal appropriate to the scientific discipline. Journals known to publish validation studies include *ATLA*, *Toxicology In Vitro* and *Environmental Health Perspectives*. The background information and all related data should be publicly accessible. The Group stressed that publication of the validation study results in a peer review journal is not equivalent to, and does not replace, peer review of the validation study.

18. A maximum of information on planned, ongoing, or completed validation studies should be made accessible to the public as early as possible in the appropriate ways available. All appropriate background information and all related data should be publicly accessible, and the raw data should be kept for a reasonable time.

19. Public involvement is an advantage and should be encouraged. The validation and regulatory acceptance processes and decision making are best organized according to the cultural and legal background of the region.

20. The Group would be willing to have the ICCVAM principles appended to the document as an example.

Peer review

21. The sections of the draft document that describe the peer review process should be better integrated.

22. Peer review is necessary. There are many ways to manage a study and to peer-review a study. These concepts are muddied in the draft Guidance Document and should be clarified. There is a distinction between the management of the validation study and the management of the peer review of that study, and the validation process must be separate from the peer review of the validation. The identification and selection of the peer-review panel must be balanced and the process must be transparent. However, the management activities of the VMG do not have to be transparent.

23. With regard to paragraph 93, the Group supported the concept that when an organization is selected to manage the peer review, it should be recognized as independent, and scientifically unbiased, in addition to being reputable, credible, and competent. Examples of such organisations include ICCVAM, ECVAM, national science academies, etc. It was noted that the two current validation management centers, ECVAM and ICCVAM, are working toward mutual recognition so as to avoid duplication of validation and/or peer review efforts. paragraph 93 contains too much detail on how the peer review should be practically conducted. What should be stated is that the panel the should communicate as appropriate.

24. The focus of paragraph 94 needs to be changed. The peer review panel needs to reach a conclusion. Ideally a consensus should be reached but, if not, the positions of those who dissent or abstain must be included in the report. Not all of the factors mentioned for consensus in the paragraph are appropriate.

25. The center that was involved in the management of the validation study could participate in the peer review of the study if it contributes essential expertise and would represent only a minority of the membership of the review panel.

26. Paragraph 95 needs significant modification to make clear the expertise needs to exist within the panel as a whole, and not within each individual member. The total panel should have expertise in the appropriate disciplines for the study under review for instance, validation, technical aspects, statistics, clinical science, general toxicology, etc. The last bullet point in paragraph 95 should be deleted or rewritten. The validation process is a scientific exercise and should not be designed to be compatible with a regulatory guideline.The question was discussed of whether the committee as a whole must be independent, or whether it was sufficient that the members be independent. To ensure independence the interests of the participants should be clearly stated and transparent. The Group concluded that the members should be independent and that the panel as a whole have credibility as an unbiased, scientific group. If additional expertise is needed, non-independent experts can be used as advisors.

27. There was general agreement that a standing peer-review committee of nominated government representatives cannot be considered as independent, because it is composed of individuals who are required to follow their governments position. Another problem with a standing committee is that it may not have the requisite expertise for the questions it has to address. Members of a regulatory organisation

could participate in the peer review in cases where they could contribute essential expertise, and would represent only a minority of the membership of the review panel.

28. Government representatives can be appropriate members of the peer review panel if they have significant expertise but only if they do not represent the views of their governments but act as individual scientists. This is true for members of any group who are on the panel.

29. Peer involvement (peer consultation) cannot substitute for peer review because they are completely separate processes. Peer review is essential; peer involvement may provide information and resources, but is not essential to validation.

Recommended editorial changes:

- The title for Section V.2 should be "Composition of the Peer Review Panel" not "Selection of Peer Reviewers."
- Paragraph 38 should clearly separate the concepts and practices of prevalidation and validation. The document needs better definitions of optimization and prevalidation.
- The Group agreed with paragraph 92, but added "with at least equivalent expertise". The footnote to paragraph 92 should be revised so that it does not blur the distinction between the peer review and validation management groups.
- The last bullet point in paragraph 95 should be deleted. The validation process is a scientific exercise and should not be designed to be compatible with a regulatory guideline.

Table 1. Comparison of the definitions and roles of the various validation bodies in Europe Japan, and the US[2]

Workstep	Europe	Japan	US
Test Development	Industry, academia, ECVAM, etc.	Similar; academia, industry, NGO, government	Similar
Validation (Study) Management Teams	The group that manages a particular validation study and produces a report. In some cases ECVAM contracts with a specific team to do this work.	Industry usually	ICCVAM committee with NICEATM. This group puts all the information together for the peer review and prepares the package that the peer review committee will review.
Peer Review Processor	In most cases the validation management team prepares the report and a statement that says whether the method is valid and the conditions under which it is valid. The processor gives this package to the peer review committee. In other cases another independent group may put together multiple independent reports into a package and prepare the statement for the peer review panel.	Same as the validation study management team.	ICCVAM committee with NICEATM. This group puts all the information together for the peer review and prepares the package that the peer review committee will review.
Peer consultation	People with specific expertise, e.g., academic or regulatory, are invited to be involved in the test development and validation, and provide expertise necessary to get the test method validated to meet the needs of regulatory agencies and others who would use the test	May have a scientific advisory team during this process.	The ICCVAM committee could act here.
Peer Review	An independent group that evaluates the statement made by the peer review processor or sponsor as to whether the method is valid.	*Ad hoc* team nominated by the government. This can also be a government advisory panel which is a standing committee.	Ad hoc team nominated by ICCVAM.

2. Although discussed in plenary, time did not allow to reach consensus on the table.